ISSN 1066-5307, Mathematical Methods of Statistics, 2015, Vol. 24, No. 3, pp. 163–199. © Allerton Press, Inc., 2015.

# Multivariate Wavelet Density and Regression Estimators for Stationary and Ergodic Continuous Time Processes: Asymptotic results

**S. Bouzebda[1*], S. Didi[2**], and L. El Hajj[2***]**

[1]*Sorbonne Univ., Univ. de Technol. de Compiègne, Compiègne, France*

[2]*Univ. Pierre et Marie Curie, Paris, France*

Received January 6, 2015; in final form, June 26, 2015

**Abstract**—In the present paper, we are mainly concerned with the nonparametric estimation of the density as well as the regression function, related to stationary and ergodic continuous time processes, by using orthonormal wavelet bases. We provide the strong uniform consistency properties with rates of these estimators, over compact subsets of $\mathbb{R}^d$, under a general ergodic condition on the underlying processes. We characterize the asymptotic normality of considered wavelet-based estimators under easily verifiable conditions. The asymptotic properties of these estimators are obtained by means of the martingale approach.

**Keywords:** multivariate regression estimation, multivariate density estimation, stationarity, ergodicity, rates of strong convergence, wavelet-based estimators, martingale differences, continuous time processes.

**2000 Mathematics Subject Classification:** 62605; 60F15; 62M10.

**DOI:** 10.3103/S1066530715030011

## 1. INTRODUCTION

Let $\mathbf{X} = \{\mathbf{X}_t : t \geq 0\}$ be a $\mathbb{R}^d$-valued stationary process having a common density function $f(\cdot)$ with respect to the Lebesgue measure. On the basis of the sample $\{\mathbf{X}_t : 0 \leq t \leq T\}$ observed from the process $\mathbf{X}$, the kernel estimate of the density function $f(\cdot)$ is defined (see, e.g., Banon (1978)), for any $\mathbf{x} \in \mathbb{R}^d$, by

$$f_T(\mathbf{x}) = \frac{1}{Th_T^d} \int_0^T K\left(\frac{\mathbf{x} - \mathbf{X}_t}{h_T}\right) dt,$$

where the kernel $K(\cdot)$ is any function which satisfies some regularity conditions and $(h_T)_{T \geq 0}$ is a sequence of positive constants converging to zero and $Th_T^d \to \infty$ as $T \to \infty$. In the traditional kernel methods for curve estimation, it has been widely regarded that the performance of the kernel methods depends largely on the smoothing bandwidth, and depends very little on the form of the kernel. Most kernels used are symmetric kernels and, once chosen, are fixed. This may be efficient for estimating curves with unbounded support, but not for curves which have compact support or subset of the whole real line and are discontinuous at boundary points. For curves of this type, conventional kernel or orthogonal-series techniques are not adequate and cause boundary bias which is quite difficult to remove. In such situations, wavelet methods perform relatively well. For finer local analysis and good asymptotic properties the wavelet estimator is certainly the method to be chosen against kernel method estimation. A great advantage of the wavelet methods in statistics is to provide adaptive procedures in the sense that they automatically adapt to the regularity of the object to be estimated. Another advantage

[*]E-mail: salim.bouzebda@upmc.fr
[**]E-mail: sultana.didi@etu.upmc.fr
[***]E-mail: layal.elhajj@gmail.com

of the wavelet procedures is their remarkable facility of use. For the general theory of wavelets we refer to Meyer (1992), Daubechies (1992), Mallat (2009) and Vidakovic (1999) among others. The use of wavelets in various curve estimation problems is surveyed in Härdle *et al.* (1998), where approximation properties of wavelets are discussed in detail. For recent references on the subject refer to Giné and Madych (2014), Li (2014), Bouzebda and Didi (2015), and Giné and Nickl (2009). Leblanc (1995) established the $L_2$ error of the linear wavelet estimator of $f(\cdot)$ in the univariate setting, which converges with the rate $1/\sqrt{T}$ when $f(\cdot)$ is in a Besov space and the underlying random process $\mathbf{X} = \{\mathbf{X}_t \colon t \geq 0\}$ is assumed to be strongly mixing. Also Masry (1997, 2000)'s seminal papers, studied the rates of strong convergence for wavelet-based estimation of the density and the regression functions, which are uniform over compact subsets of $\mathbb{R}^d$. In the present work, we do not assume anything beyond ergodicity of the underlying process. It is worth noticing that strong mixing implies ergodicity; see, e.g., Remark 2.6 on p. 50 in combination with Proposition 2.8 on p. 51 in Bradley (2007). Hence the present work extends the scope of applications provided by the existing works. On the other hand, we mention that there exist interesting processes which are ergodic but not mixing. Andrews (1984) has shown that a stationary AR(1) process $(X_t)_{t\in\mathbb{Z}}$ obeying

$$X_t = \theta X_{t-1} + \varepsilon_t$$

with i.i.d. Bernoulli distributed innovations is not strongly mixing. However, ergodicity is preserved under taking functions of an ergodic process. If $(\varepsilon_t)_{t\in\mathbb{Z}}$ is a strictly stationary ergodic process and

$$Y_t = \vartheta\big((\ldots, \varepsilon_{t-1}, \varepsilon_t), (\varepsilon_{t+1}, \varepsilon_{t+2}, \ldots)\big)$$

for some Borel-measurable function $\vartheta(\cdot)$, then $(Y_t)_{t\in\mathbb{Z}}$ is also ergodic; see Proposition 2.10 on p. 54 in Bradley (2007). Since the above autoregressive process can be represented as a linear process in the $\varepsilon_t$'s, it follows that it is also ergodic. Another example of an ergodic and non-mixing process is considered in Section 5.3 of Leucht and Neumann (2013). Indeed, assume that the process $\{(T_i, \lambda_i) \colon i \in \mathbb{Z}\}$ is strictly stationary with $T_i \mid \mathcal{T}_{i-1} \sim \text{Poisson}(\lambda_i)$, $\mathcal{T}_i$ being the $\sigma$-field generated by $(T_i, \lambda_i, T_{i-1}, \lambda_{i-1}, \ldots)$. We assume that $\lambda_i = \kappa(\lambda_{i-1}, T_{i-1})$, where

$$\kappa \colon [0, \infty) \times \mathbb{N} \to (0, \infty).$$

However, this process is not mixing in general; see Remark 3 of Neumann (2011) for a counterexample. We refer to Leucht and Neumann (2013) for further details and motivations for the use of ergodicity assumption. One of their arguments is that for certain classes of processes, it can be much easier to prove ergodicity rather than mixing. It is known that any sequence $(\varepsilon_t)_{t\in\mathbb{Z}}$ of i.i.d. random variables is ergodic. Hence, it is immediately clear that $(Y_t)_{t\in\mathbb{Z}}$ as above is also ergodic. It is worth noticing that the ergodicity is implied by all mixing conditions, being weaker than all of them. This hypothesis seems to be the most naturally adapted and provides a better framework to study data series, for example, generated by noisy chaos.

To the best of our knowledge, the results presented here, respond to a problem that has not been studied systematically until present, and it gives the main motivation to this paper.

The paper is organized as follows. General notation and definitions of the multiresolution analysis are given in Section 2. The assumptions and asymptotic properties of the wavelet-based density estimators are given in Section 3, which includes the optimal uniform convergence rates. Section 4 is devoted to wavelet-based estimation of the regression function. We establish the uniform convergence rates and characterize the asymptotic normality under the ergodicity condition in Section 4.1. Some concluding remarks and possible future developments are mentioned in Section 5. To avoid interrupting the flow of the presentation, all mathematical developments are relegated to Section 6.

## 2. MULTIRESOLUTION ANALYSIS

In this section, we set out some definitions and notation for later use. A general introduction to the theory of wavelets can be found in Meyer (1992), Daubechies (1992), Mallat (2009), and Vidakovic (1999). Following Meyer (1992) a multiresolution analysis on the Euclidean space $\mathbb{R}^d$ is a decomposition of the space $L_2(\mathbb{R}^d)$ into an increasing sequence of closed subspaces $\{V_j \colon j \in \mathbb{Z}\}$ such that

(i) $V_j \subset V_{j+1}, \quad j \in \mathbb{Z}$,

(ii) $\cap_j V_j = \{0\}$, $\overline{\cup_j V_j} = L_2(\mathbb{R}^d)$,

(iii) $f(\mathbf{x}) \in V_j \Leftrightarrow f(2\mathbf{x}) \in V_{j-1}$, $f(x) \in V_j \Rightarrow f(\mathbf{x}+\mathbf{k}) \in V_j$, $\mathbf{k} \in \mathbb{Z}^d$,

where $V_0$ is closed under integer translation. Finally, there exists a scale function $\phi(\cdot) \in L_2(\mathbb{R}^d)$ with

$$\int_{\mathbb{R}^d} \phi(\mathbf{x})\, d\mathbf{x} = 1,$$

such that $\{\phi_{\mathbf{k}}(\mathbf{x}) = \phi(\mathbf{x}-\mathbf{k})\colon \mathbf{k} \in \mathbb{Z}^d\}$ is an orthonormal basis for $V_0$. It follows that $\{\phi_{j,\mathbf{k}}(\mathbf{x}) = 2^{jd/2}\phi_j(2^j\mathbf{x}-\mathbf{k})\colon \mathbf{k} \in \mathbb{Z}^d\}$ is an orthonormal basis for $V_j$. The multiresolution analysis is called $r$-regular if $\phi(\cdot) \in C^r$ and all its partial derivatives up to total order $r$ are rapidly decreasing, i.e., for every integer $i > 0$, there exists a constant $A_i$ such that

$$|(D^{\boldsymbol{\beta}}\phi)(\mathbf{x})| \le \frac{A_i}{(1+\|\mathbf{x}\|)^i} \qquad \text{for all} \quad |\boldsymbol{\beta}| \le r, \tag{2.1}$$

where

$$(D^{\boldsymbol{\beta}}\phi)(\mathbf{x}) = \frac{\partial^{\boldsymbol{\beta}}\phi(\mathbf{x})}{\partial^{\beta_1}x_1 \cdots \partial^{\beta_d}x_d} \qquad \text{and} \qquad \boldsymbol{\beta} = (\beta_1, \ldots, \beta_d), \quad |\boldsymbol{\beta}| = \sum_{i=1}^{d} \beta_i.$$

Throughout the sequel, the multiresolution is assumed to be $r$-regular. If $W_j$ denotes the orthogonal complement of $V_j$ in $V_{j+1}$, i.e.,

$$V_j \oplus W_j = V_{j+1},$$

then $L_2(\mathbb{R}^d)$ can be decomposed as

$$L_2(\mathbb{R}^d) = \bigoplus_{j\in\mathbb{Z}} W_j, \tag{2.2}$$

or, equivalently, as

$$L_2(\mathbb{R}) = V_{j_0} \oplus \bigoplus_{j\ge j_0} W_j. \tag{2.3}$$

Then there exist $N = 2^d - 1$ wavelet functions $\{\psi_i(\mathbf{x}), i = 1, \ldots, N\}$ such that

(W.1) $\{\psi_i(\mathbf{x}-\mathbf{k})\colon \mathbf{k} \in \mathbb{Z}^d, i = 1, \ldots, N\}$ is an orthonormal basis for $W_0$,

(W.2) with $\psi_{i,j,\mathbf{k}}(\mathbf{x}) = 2^{jd/2}\psi_i(2^j\mathbf{x}-\mathbf{k})$ the functions $\{\psi_{i,j,\mathbf{k}}(\mathbf{x})\colon i = 1, \ldots, N, \mathbf{k} \in \mathbb{Z}^d, j \in \mathbf{Z}\}$ constitute an orthonormal basis for $L_2(\mathbb{R}^d)$,

(W.3) $\psi_i(\cdot)$ has the same regularity as $\phi(\cdot)$ and both functions have compact support $[-L, L]^d$ for some $L > 0$.

For any $f(\cdot) \in L_2(\mathbb{R}^d)$ we have the orthonormal representation, for any integer $m$,

$$f(\mathbf{x}) = \sum_{\mathbf{k}\in\mathbb{Z}^d} a_{m\mathbf{k}}\phi_{m,\mathbf{k}}(\mathbf{x}) + \sum_{j\ge m}\sum_{i=1}^{N}\sum_{\mathbf{k}\in\mathbb{Z}^d} b_{ij\mathbf{k}}\psi_{i,j,\mathbf{k}}(\mathbf{x}), \tag{2.4}$$

where

$$a_{m\mathbf{k}} = \int_{\mathbb{R}^d} f(\mathbf{u})\phi_{m,\mathbf{k}}(\mathbf{u})\, d\mathbf{u}$$

and

$$b_{i,j,\mathbf{k}} = \int_{\mathbb{R}^d} f(\mathbf{u})\psi_{i,j,\mathbf{k}}(\mathbf{u})\, d\mathbf{u}.$$

Note that the orthogonal projection of $f(\cdot)$ on $V_l$ can be written in two equivalent ways, for any $m \leq l$,

$$(P_{V_l}f)(\mathbf{x}) := \sum_{\mathbf{k}\in\mathbb{Z}^d} a_{l\mathbf{k}}\phi_{l,\mathbf{k}}(\mathbf{x}) = \sum_{\mathbf{k}\in\mathbb{Z}^d} a_{m\mathbf{k}}\phi_{m,\mathbf{k}}(\mathbf{x}) + \sum_{j=m}^{l}\sum_{i=1}^{N}\sum_{\mathbf{k}\in\mathbb{Z}^d} b_{ij\mathbf{k}}\psi_{i,j,\mathbf{k}}(\mathbf{x}). \tag{2.5}$$

## 3. MULTIVARIATE DENSITY ESTIMATION

Throughout the sequel, assume that the density function $f(\cdot) \in L_2(\mathbb{R}^d)$. Then $f(\cdot)$ admits the wavelet representation (2.4). Given the sample $\{\mathbf{X}_t\colon 0 \leq t \leq T\}$ we estimate the coefficients $\{a_{m\mathbf{k}}\}$ and $\{b_{ij\mathbf{k}}\}$ by

$$\widehat{a}_{m\mathbf{k}} = \frac{1}{T}\int_0^T \phi_{m,\mathbf{k}}(\mathbf{X}_t)\,dt \qquad \text{and} \qquad \widehat{b}_{ij\mathbf{k}} = \frac{1}{T}\int_0^T \psi_{i,j,\mathbf{k}}(\mathbf{X}_t)\,dt, \tag{3.1}$$

and note that these estimates are unbiased, that is

$$\mathbb{E}(\widehat{a}_{m\mathbf{k}}) = a_{m\mathbf{k}} \qquad \text{and} \qquad \mathbb{E}(\widehat{b}_{ij\mathbf{k}}) = b_{ij\mathbf{k}}.$$

A linear estimate of $f(\cdot)$ can be obtained from (2.4) by

$$\widehat{f}_T(\mathbf{x}) = \sum_{\mathbf{k}\in\mathbb{Z}^d} \widehat{a}_{m\mathbf{k}}\phi_{m,\mathbf{k}}(\mathbf{x}) \tag{3.2}$$

or, equivalently, as

$$\widehat{f}_T(\mathbf{x}) = \sum_{\mathbf{k}\in\mathbb{Z}^d} \widehat{a}_{j_0\mathbf{k}}\phi_{j_0,\mathbf{k}}(\mathbf{x}) + \sum_{j=j_0}^{m}\sum_{i=1}^{N}\sum_{\mathbf{k}\in\mathbb{Z}^d} \widehat{b}_{ij\mathbf{k}}\psi_{i,j,\mathbf{k}}(\mathbf{x}), \tag{3.3}$$

for any $j_0 \leq m$. Here the resolution level $m = m(T) \to \infty$ at a rate specified below. We assume that $\phi(\cdot)$ and $\psi_i(\cdot)$ have a compact support so that the summations above are finite for each fixed $\mathbf{x}$ (note that in this case the support of $\phi(\cdot)$ and $\psi_i(\cdot)$ is a monotonically increasing function of their degree of differentiability, see Daubechies (1992)). We focus our attention on multivariate linear estimators (3.2) and (3.3) which will be shown to have uniform almost sure convergence rates over compact sets.

We will denote by $\mathcal{F}_t$ the $\sigma$-field generated by $\{\mathbf{X}_s\colon 0 \leq s \leq t\}$. For a small constant $\delta > 0$, define by $f_{\mathbf{X}_t}^{\mathcal{F}_{t-\delta}}(\cdot)$ the conditional density of $\mathbf{X}_t$ given the $\sigma$-field $\mathcal{F}_{t-\delta}$.

The following assumptions will be needed throughout the paper.

(C.1) For any $\mathbf{x} \in \mathbb{R}^d$

$$\lim_{T\to\infty}\sup_{\mathbf{x}\in\mathbb{R}^\mathbf{d}}\left|\frac{1}{T}\int_0^T f^{\mathcal{F}_{t-\delta}}(\mathbf{x})\,dt - f(\mathbf{x})\right| = 0, \qquad \text{in the } a.s. \text{ and } L^2 \text{ sense.}$$

At this point, we may refer to Peškir (1998) for further details.

(C.2) The density $f(\cdot)$ is continuous and has bounded partial derivatives of order $r$, that is, there exists a constant $0 < \mathfrak{C} < \infty$ such that

$$\sup_{\mathbf{x}\in D}\left|\frac{\partial^r f(\mathbf{x})}{\partial x_1^{k_1}\ldots\partial x_d^{k_d}}\right| \leq \mathfrak{C}, \qquad k_1,\ldots,k_d \geq 0, \quad 0 < k_1 + \cdots + k_d = r.$$

Define the kernel $K(\mathbf{u},\mathbf{v})$ by

$$K(\mathbf{u},\mathbf{v}) := \sum_{\mathbf{k}\in\mathbb{Z}^d}\phi(\mathbf{u}-\mathbf{k})\phi(\mathbf{v}-\mathbf{k}) \quad \text{and} \quad h_n = 2^{-m(n)}. \tag{3.4}$$

**Theorem 3.1.** *Assume that*

$$m(T) = m \to \infty \qquad \textit{and} \qquad \frac{2^{dm(T)}\log T}{T} \to 0 \quad \textit{as}\ \ T \to \infty.$$

*For every compact subset $D \subset \mathbb{R}^d$, and under assumption* (C.1)*, we have almost surely*

$$\sup_{\mathbf{x}\in D}\big|\widehat{f}_T(\mathbf{x}) - \mathbb{E}(\widehat{f}_T(\mathbf{x}))\big| = O\bigg(\bigg(\frac{(\log T)2^{dm(T)}}{T}\bigg)^{1/2}\bigg) + O(2^{-dm(T)/2}).$$

The proof of Theorem 3.1 is presented in Section 6.

**Theorem 3.2.** *Assume that*

$$m(T) = m \to \infty \qquad \textit{and} \qquad \frac{2^{dm(T)}\log T}{T} \to 0 \quad \textit{as}\ \ T \to \infty.$$

*For every compact subset $D \subset \mathbb{R}^d$, and under assumptions* (C.1) *and* (C.2)*, we have almost surely*

$$\sup_{\mathbf{x}\in D}|\widehat{f}_T(\mathbf{x}) - f(\mathbf{x})| = O\bigg(\bigg(\frac{(\log T)2^{dm(T)}}{T}\bigg)^{1/2}\bigg) + O(2^{-dm(T)/2}) + O(2^{-drm(T)}).$$

The proof of Theorem 3.2 is presented in Section 6.

## 4. MULTIVARIATE REGRESSION ESTIMATION

Let $\{\mathbf{X}_i, Y_i\}$ be jointly stationary processes and $\varphi(\cdot)$ be a Borel measurable function on the real line. Assume that $\mathbb{E}[|\varphi(Y_1)|] < \infty$ and define the regression function

$$m(\mathbf{x}, \varphi) = \mathbb{E}[\varphi(Y_1) \mid \mathbf{X}_1 = \mathbf{x}].$$

The introduction of the function $\varphi(\cdot)$, as was pointed out in the Introduction, allows us to include some important special cases:

- $\varphi(Y) = \mathbf{1}\{Y \le y\}$ gives the conditional distribution of $Y_1$ given $\mathbf{X}_1 = \mathbf{x}$.
- $\varphi(Y) = Y^k$ gives the conditional moments of $Y_1$ given $\mathbf{X}_1 = \mathbf{x}$.

Recall that the probability density $f(\mathbf{x}) = f_{\mathbf{X}_1}(\mathbf{x})$ of $\mathbf{X}_1$ is assumed to exist and be bounded and suppose in addition that

$$\mathbb{E}[|\varphi(Y_1)|^p] < \infty \qquad \text{for} \quad p \ge 1.$$

Now define

$$H(\mathbf{x}) = m(\mathbf{x}, \varphi)f(\mathbf{x}).$$

Notice that $H(\cdot) \in L_p(\mathbb{R}^d)$. It follows that for $p = 2$, $H(\cdot)$ has the $L_2$ orthonormal representation

$$H(\mathbf{x}) = \sum_{\mathbf{k}\in\mathbb{Z}^d} a'_{m\mathbf{k}}\phi_{m,\mathbf{k}}(\mathbf{x}) + \sum_{j\ge m}\sum_{i=1}^{N}\sum_{\mathbf{k}\in\mathbb{Z}^d} b'_{ij\mathbf{k}}\psi_{i,j,\mathbf{k}}(\mathbf{x}), \tag{4.1}$$

where

$$a'_{m\mathbf{k}} = \int_{\mathbb{R}^d} H(\mathbf{u})\phi_{m,\mathbf{k}}(\mathbf{u})d\mathbf{u}, \qquad b'_{i,j,\mathbf{k}} = \int_{\mathbb{R}^d} H(\mathbf{u})\psi_{i,j,\mathbf{k}}(\mathbf{u})d\mathbf{u}.$$

Suppose now that we observe a sequence $\{\mathbf{X}_i, Y_i\}_{i=1}^n$ of copies of $(\mathbf{X}, Y)$ that is assumed to be stationary and ergodic. Given the preceding notation, we consider the estimates of the coefficients $\{a'_{m\mathbf{k}}\}$ and $\{b'_{ij\mathbf{k}}\}$ given by

$$\widehat{a}'_{m\mathbf{k}} = \frac{1}{T}\int_0^T \varphi(Y_t)\phi_{m,\mathbf{k}}(\mathbf{X}_t)\,dt \qquad \text{and} \qquad \widehat{b}'_{ij\mathbf{k}} = \frac{1}{T}\int_0^T \varphi(Y_t)\psi_{i,j,\mathbf{k}}(\mathbf{X}_t)\,dt. \tag{4.2}$$

A linear estimate of $H(\cdot)$ can be obtained, from (4.1), by

$$\widehat{H}_T(\mathbf{x}) = \sum_{\mathbf{k}\in\mathbb{Z}^d} \widehat{a}'_{m\mathbf{k}}\phi_{m,\mathbf{k}}(\mathbf{x}), \tag{4.3}$$

or, equivalently, as

$$\widehat{H}_T(\mathbf{x}) = \sum_{\mathbf{k}\in\mathbb{Z}^d} \widehat{a}'_{j_0\mathbf{k}}\phi_{j_0,\mathbf{k}}(\mathbf{x}) + \sum_{j=j_0}^{m}\sum_{i=1}^{N}\sum_{\mathbf{k}\in\mathbb{Z}^d} \widehat{b}'_{ij\mathbf{k}}\psi_{i,j,\mathbf{k}}(\mathbf{x}), \tag{4.4}$$

for any $j_0 \le m$. Now define the linear estimates of the regression function $m(\mathbf{x},\varphi)$ by

$$m_T(\mathbf{x},\varphi) := \frac{\widehat{H}_T(\mathbf{x},\varphi)}{\widehat{f}_T(\mathbf{x})} := \frac{\widehat{H}_T(\mathbf{x})}{\widehat{f}_T(\mathbf{x})}. \tag{4.5}$$

The estimates (4.5) can be viewed as wavelet-based Nadaraya–Watson estimates of the regression function $m(\mathbf{x},\varphi)$. This family of estimators was deeply investigated by Masry (2000). Let us denote by $g(\cdot)$ the density function of $(\mathbf{X},Y)$, and by $\rho$ the density function of $Y$. Let $\mathcal{G}_t$ be the $\sigma$-field generated by $\{(\mathbf{X}_s,Y_s)\colon 0\le s\le t\}$, and for $\delta>0$ small enough, let $g^{\mathcal{G}_{t-\delta}}(\cdot)$ and $\rho^{\mathcal{G}_{t-\delta}}(\cdot)$ be the conditional densities of $(\mathbf{X},Y)$ and $Y$ respectively, given the $\sigma$-field $\mathcal{G}_{t-\delta}$. Define the $\sigma$-field

$$\mathcal{S}_{t,\delta} = \sigma\big((\mathbf{X}_s,Y_s);(\mathbf{X}_r)\colon 0\le s\le t, t\le r\le t+\delta\big).$$

We need the following assumptions.

(N.0) $\mathbb{E}\big[|\varphi(Y_0)|^{\nu}\big] < \infty$ for some $\nu \ge 1$.

(N.1) For any $y\in\mathbb{R}$

$$\lim_{T\to\infty}\sup_{y\in\mathbb{R}}\left|\frac{1}{T\rho(y)}\int_0^T \rho^{\mathcal{G}_{t-\delta}}(y)\,dt - 1\right| = 0 \qquad \text{in the } a.s. \text{ and } L^2 \text{ sense.}$$

We may refer to Peškir (1998) for further details.

Our main results concerning the strong consistency with rate of $\widehat{H}_T(\mathbf{x})$ are given in the following theorems.

**Theorem 4.1.** *Assume that*

$$m(T) = m\to\infty \qquad \text{and} \qquad \frac{2^{dm(T)}\log T}{T}\to 0 \quad \text{as } T\to\infty.$$

*Let $L_T$ be a sequence of numbers such that*

$$L_T^p 2^{-m(T)(d(p-1)+p)} = O(1) \quad \text{as } T\to\infty.$$

*For every compact subset $D\subset\mathbb{R}^d$, and under the assumptions* (N.0)–(N.1)*, we have almost surely*

$$\sup_{\mathbf{x}\in D}|\widehat{H}_T(\mathbf{x}) - \overline{H}_T(\mathbf{x})| = O\bigg(\bigg(\frac{(\log T)2^{dm(T)}}{T}\bigg)^{1/2}\bigg) + O(2^{-dm(T)/2}),$$

*where*

$$\overline{H}_T(\mathbf{x}) = \frac{1}{Th_T^d}\int_0^T \mathbb{E}\Big[\varphi(Y_t)K\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{X}_t}{h_T}\Big)\,\Big|\,\mathcal{G}_{t-\delta}\Big]\,dt.$$

The proof of Theorem 4.1 is presented in Section 6.

The following additional assumptions are needed for our second strong consistency result.

(N.2) The regression function $m(\cdot)$ is continuous and has bounded partial derivatives of order $r$, that is, there exists a constant $0 < \mathfrak{C}' < \infty$ such that

$$\sup_{\mathbf{x}\in D}\left|\frac{\partial^r m(\mathbf{x})}{\partial x_1^{k_1}\dots\partial x_d^{k_d}}\right| \le \mathfrak{C}', \qquad k_1,\dots,k_d \ge 0, \quad k_1+\dots+k_d = r;$$

(N.3) For any $(\mathbf{x}, y) \in \mathbb{R}^{d+1}$ and any $\delta > 0$ small enough,

$$\lim_{n\to\infty}\sup_{(\mathbf{x},y)\in\mathbb{R}^{d+1}}\left|\frac{1}{T}\int_0^n g^{\mathcal{G}_{t-\delta}}(\mathbf{x}, y)\,dt - g(\mathbf{x}, y)\right| = 0 \qquad \text{in the } a.s. \text{ and } L^2 \text{ sense.}$$

(N.4) For any $\mathbf{x} \in \mathbb{R}^d$, there are two positive constants $\lambda$ and $\Lambda$ such that

(i) $\inf_{\mathbf{x}\in D}|f(\mathbf{x})| > \lambda > 0$,

(ii) $\sup_{\mathbf{x}\in D}|m(\mathbf{x}, \varphi)| \le \Lambda < \infty$.

**Comments on hypotheses.** Condition (N.1) implies the ergodic nature of the data as given, for instance, in Györfi *et al.* (1989), assuming that $\rho^{\mathcal{G}_{t-\delta}}$ and $g^{\mathcal{G}_{t-\delta}}$ belong to the space $\mathcal{C}^0$, at least, of continuous functions, which is a separable Banach space. Moreover, approximating the integrals

$$\int_0^T \rho^{\mathcal{G}_{t-\delta}}(y)\,dt \qquad \text{and} \qquad \int_0^T g^{\mathcal{G}_{t-\delta}}(\mathbf{x}, y)\,dt$$

by their Riemann's sums, it follows that

$$T^{-1}\int_0^T \rho^{\mathcal{G}_{t-\delta}}(y)\,dt \simeq n^{-1}\sum_{i=1}^n \rho^{\mathcal{G}_{T_i-\delta}}(y) = n^{-1}\sum_{j=1}^n \rho^{\mathcal{G}_{(j-1)\delta}}(y)$$

and

$$T^{-1}\int_0^T g^{\mathcal{G}_{t-\delta}}(\mathbf{x}, y)\,dt \simeq n^{-1}\sum_{i=1}^n g^{\mathcal{G}_{T_i-\delta}}(\mathbf{x}, y) = n^{-1}\sum_{j=1}^n g^{\mathcal{G}_{(j-1)\delta}}(\mathbf{x}, y).$$

Since the processes $(\mathbf{X}_{T_j}, Y_{T_j})_{j\ge1}$ and $(Y_{T_j})_{j\ge1}$ are stationary and ergodic (see Proposition 4.3 of Krengel (1985)) following Delecroix (1987) (see Lemma 4 and Corollary 1 together with their proofs), one may prove that the sequences $(\rho^{\mathcal{G}_{(j-1)\delta}}(y))_{j\ge1}$ and $(g^{\mathcal{G}_{(j-1)\delta}}(\mathbf{x}, y))_{j\ge1}$ of conditional densities are stationary and ergodic. Moreover, making use of Beck (1963)'s theorem (see, for instance, Györfi *et al.* (1989), Theorem 2.1.1), it follows that

$$\lim_{T\to\infty}\sup_{y\in\mathbb{R}}\left|\frac{1}{T}\int_0^T \rho^{\mathcal{G}_{t-\delta}}(y)\,dt - \mathbb{E}(\rho^{\mathcal{G}_{-\delta}}(y))\right| = \lim_{T\to\infty}\sup_{y\in\mathbb{R}}\left|\frac{1}{T}\int_0^T \rho^{\mathcal{G}_{t-\delta}}(y)\,dt - \rho(y)\right| = 0 \;\; a.s.$$

and

$$\lim_{T\to\infty}\sup_{\mathbf{x}\in\mathbb{R}^d}\left|\frac{1}{T}\int_0^T g^{\mathcal{G}_{t-\delta}}(\mathbf{x}, y)\,dt - \mathbb{E}(g^{\mathcal{G}_{-\delta}}(\mathbf{x}, y))\right| = \lim_{T\to\infty}\sup_{\mathbf{x}\in\mathbb{R}^d}\left|\frac{1}{T}\int_0^T g^{\mathcal{G}_{t-\delta}}(\mathbf{x}, y)\,dt - g(\mathbf{x})\right| = 0 \;\; a.s.$$

It is then clear that both Conditions (N.1) and (N.3) are satisfied.

**Theorem 4.2.** *Assume that*

$$m(T) = m \to \infty \qquad \textit{and} \qquad \frac{2^{dm(T)}\log T}{T} \to 0 \quad \textit{as } n \to \infty.$$

*Let $L_T$ be a sequence of numbers such that*

$$L_T^p 2^{-m(T)(d(p-1)+p)} = O(1) \quad \textit{as } T \to \infty.$$

*For every compact subset $D \subset \mathbb{R}^d$, and if the hypotheses* (C.2), (N.0)–(N.3) *are satisfied, we have almost surely*

$$\sup_{\mathbf{x}\in D} |\widehat{H}_n(\mathbf{x}) - H(\mathbf{x})| = O\left(\left(\frac{(\log n)2^{dm(n)}}{n}\right)^{1/2}\right) + O(2^{-dm(n)/2}) + O(2^{-drm(T)}).$$

The proof of Theorem 4.2 is presented in Section 6.

Our main result concerning the strong consistency with rate of $m_n(\mathbf{x}, \varphi)$ is given in the following theorem.

**Theorem 4.3.** *Assume that*

$$m(T) = m \to \infty \qquad and \qquad \frac{2^{dm(T)} \log T}{T} \to 0 \quad as \ T \to \infty.$$

*Let $L_T$ be a sequence of numbers such that*

$$L_T^p 2^{-m(T)(d(p-1)+p)} = O(1) \; as \;\; T \to \infty.$$

*For every compact subset $D \subset \mathbb{R}^d$, and if the hypothesis* (C.1), (C.2), (N.0)–(N.4) *are satisfied, we have almost surely*

$$\sup_{\mathbf{x}\in D} |m_T(\mathbf{x}, \varphi) - m(\mathbf{x}, \varphi)| = O\left(\left(\frac{(\log T)2^{dm(T)}}{T}\right)^{1/2}\right) + O(2^{-dm(T)/2}) + O(2^{-drm(T)}).$$

The proof of Theorem 4.3 is presented in Section 6.

**Some comments on the results.** In deriving results for nonparamptric estimators, in particular, the wavelet estimator, we generally decompose the error in two terms: the stochastic component and the deterministic one, i.e., the bias. For bounding the stochastic part, i.e., in order to find almost sure bounds on the stochastic process, we especially need assumptions about the wavelet functions $\{\psi_i(\mathbf{x}), i = 1, \ldots, N\}$ in addition to general assumptions about the density or reversion. The deterministic component requires more assumptions about the smoothness of $m(\cdot, \varphi)$ and the density function $f$. In the classical kernel density estimation, the limiting behavior of kernel estimator $f_n(\cdot)$, for appropriate choices of the bandwidth $a_n$, has been studied by a large number of statisticians over many decades. For good sources of references to research literature in this area along with statistical applications consult Devroye and Lugosi (2001), Devroye and Györfi (1985), Bosq and Lecoutre (1987), Scott (1992), Wand and Jones (1995) and Prakasa Rao (1983). In particular, the condition that $a_n \to 0$ together with $na_n^d \to \infty$ is necessary and sufficient for the convergence in probability of $f_n(\mathbf{x})$ towards the limit $f(\mathbf{x})$, independently of $\mathbf{x} \in \mathbb{R}^d$ and the density $f(\cdot)$. Theorem 3.1 (resp. Theorem 4.1) shows an intermediate result by obtaining a rate of convergence for the density estimator (resp. $\widehat{H}_T(\mathbf{x})$). Secondly, we have shown that the density estimate (resp. $\widehat{H}_T(\mathbf{x})$) converges to the true density function (resp. $H(\mathbf{x})$) with the same rate. In particular, we have proved that the difference between the right-hand side of the equation in Theorem 3.1 (resp. Theorem 4.1) and the right-hand side of the equation in Theorem 3.2 (resp. Theorem 4.2) vanishes faster than the rest of the terms. The imposed conditions on the wavelet functions permit us to exploit the smoothness of the density function or the regression function. It is well known that the best obtainable rate of convergence of the kernel estimator, in the AMISE sense, is of order $n^{-4/5}$, in the univariate case. If we discard the condition that the kernel function must be a density, the convergence rate could be faster. Indeed, the convergence rate can be made arbitrarily close to the parametric $n^{-1}$ as the order increases. In fact, Chacón *et al.* (2007) showed that the parametric rate $n^{-1}$ can be attained by the use of superkernels, and that superkernel density estimators automatically adapt to the unknown degree of smoothness of the density. The main drawback of higher-order kernels in this situation is the negative contribution of the kernel which may make the estimated density not a density itself. The interested reader may refer to, e.g., Jones *et al.* (1995), Jones and Signorini (1997) and Jones (1995). It will be of interest to give a complete and analogous discussion in our setting for wavelet type estimators.

*4.1. Asymptotic Normality*

Let us introduce

$$\sigma_\varphi^2(\mathbf{x}) = \mathrm{Var}(\varphi(Y) \mid \mathbf{X} = \mathbf{x}) = \frac{1}{f_{\mathbf{X}}(\mathbf{x})} \int_{\mathbb{R}} \{\varphi(y) - m(\mathbf{x}, \varphi)\}^2 f_{\mathbf{X},Y}(\mathbf{x}, y)\, dy, \tag{4.6}$$

and let $m_2(\varphi, \cdot)$ be the second order conditional moment of the random variable $\varphi(Y)$ defined by

$$\delta_\varphi^2(\varphi, \mathbf{x}) = \mathbb{E}[\varphi^2(Y) \mid \mathbf{X} = \mathbf{x}] = \frac{1}{f_{\mathbf{X}}(\mathbf{x})} \int_{\mathbb{R}} \varphi^2(y) f_{\mathbf{X},Y}(\mathbf{x}, y)\, dy. \tag{4.7}$$

The following additional conditions are needed for the detailed statement of our results concerning the asymptotic normality.

(N.5) The conditional mean of $Y_t$ given the $\sigma$-field $\mathcal{S}_{t-\delta,\delta}$ depends only on $\mathbf{X}_t$, i.e., for any $i \geq 1$,

$$\mathbb{E}[\varphi(Y_t) \mid \mathcal{S}_{t-\delta,\delta}] = \mathbb{E}[\varphi(Y_t) \mid \mathbf{X}_t].$$

(N.6) For any $\mathbf{x} \in \mathbb{R}^d$ and any $\delta > 0$ small enough,

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T f^{\mathcal{G}_{t-\delta}}(\mathbf{x})\, dt = f(\mathbf{x}) \qquad \text{in the } a.s. \text{ and } L^2 \text{ sense},$$

where $f^{\mathcal{G}_{t-\delta}}$ exists and continuous in a neighborhood of $\mathbf{x}$.

(N.7) For any $t \in [0, T]$ and any $\delta > 0$ small enough, and $t \leq r \leq t + \delta$,

(i) The conditional variance of $\varphi(Y_t)$ given the $\sigma$-field $\mathcal{S}_{i-1}$ depends only on $\mathbf{X}_i$, i.e., for any $i \geq 1$,

$$\mathbb{E}\big[\big(\varphi(Y_t) - m(\mathbf{x}, \varphi)\big)^2 \mid \mathcal{S}_{r,\delta}\big] = \mathbb{E}\big[\big(\varphi(Y_t) - m(\mathbf{x}, \varphi)\big)^2 \mid \mathbf{X}_t\big] = \sigma_\varphi^2(\mathbf{x}) \quad a.s.$$

(ii) The function $\sigma_\varphi^2(\mathbf{x})$ is continuous in a neighborhood of $\mathbf{x}$, that is

$$\sup_{\{\mathbf{u}\colon \|\mathbf{x}-\mathbf{u}\|<h\}} |\sigma_\varphi^2(\mathbf{u}) - \sigma_\varphi^2(\mathbf{x})| = o(1) \quad \text{as } h \to 0.$$

(N.8) The function $m_2(\varphi, \cdot)$ is a continuous function in a neighborhood of $\mathbf{x}$, that is

$$\sup_{\{\mathbf{u}\colon \|\mathbf{x}-\mathbf{u}\|<h\}} |\delta_\varphi^2(\varphi, \mathbf{u}) - \delta_\varphi^2(\varphi, \mathbf{x})| = o(1) \quad \text{as } h \to 0.$$

**Comments on hypotheses.** Assume that the random functions $f_t^{\mathcal{G}_{t-\delta}}(\mathbf{x})$, for any $t \in [0, T]$, belong to the space $\mathcal{C}^0$ of continuous functions, which is a separable Banach space. Moreover, approximating the integral $\int_0^T f_t^{\mathcal{G}_{t-\delta}}(\mathbf{x})\, dt$ by its Riemann's sum, it follows that

$$\frac{1}{T} \int_0^T f_t^{\mathcal{G}_{t-\delta}}(\mathbf{x})\, dt = \frac{1}{T} \sum_{i=1}^n \int_{T_{i-1}}^{T_i} f_t^{\mathcal{G}_{t-\delta}}(\mathbf{x})\, dt \backsimeq \frac{1}{n} \sum_{i=1}^n f_{T_{i-1}}^{\mathcal{G}_{T_{i-2}}}(\mathbf{x}).$$

Since the process $(\mathbf{X}_{T_j})_{j\geq 1}$ is stationary and ergodic, following Delecroix (1987) (see Lemma 4 and Corollary 1 together with their proofs), one may prove that the sequence $(f_{j\delta,(j-1)\delta}(\mathbf{x}))_{j\geq 1}$ of random functions is stationary and ergodic. Indeed, it suffices to replace the conditional densities in the work of Delecroix by $f_{j\delta,(j-1)\delta}$'s and the density by the function $f$.

Below, we write $Z \overset{d}{=} N(\mu, \sigma^2)$ whenever the random variable $Z$ follows a normal law with expectation $\mu$ and variance $\sigma^2$.

**Theorem 4.4.** *Assume that Conditions* (C.3), (N.0), (N.2), (N.3), (N.6), (N.7) *are fulfilled. Suppose that*

$$m(T) \to \infty, \quad T2^{-d(m(T))} \to \infty \quad as \;\; T \to \infty.$$

*We have the following convergence in distribution, for* $\mathbf{x} \in D$*, as* $T \to \infty$*,*

$$\left\{\frac{T}{2^{dm(T)}}\right\}^{1/2} \big(m_T(\mathbf{x}, \varphi) - m(\mathbf{x}, \varphi)\big) \to N\big(0, \Sigma_\varphi^2(\mathbf{x})\big),$$

*where*

$$\Sigma_\varphi^2(\mathbf{x}) = \frac{\sigma_\varphi^2(\mathbf{x})}{f_{\mathbf{X}}(\mathbf{x})} \int_{\mathbb{R}^d} \left\{\sum_{k\in\mathbb{Z}^d} \phi(\boldsymbol{k})\phi(\boldsymbol{t}+\boldsymbol{k})\right\}^2 d\mathbf{t}$$

*with* $\sigma_\varphi^2(\mathbf{x})$ *defined in* (4.6).

The proof of Theorem 4.4 is provided in Section 6.

**Theorem 4.5.** *Assume that Conditions* (N.0), (N.1), (N.2), (N.6) *and* (N.8) *are fulfilled. Suppose that*

$$m(n) \to \infty, \quad n2^{-d(m(n))} \to \infty \quad as \;\; n \to \infty.$$

*We have the following convergence in distribution, for* $\mathbf{x} \in D$ *as* $n \to \infty$*,*

$$\left\{\frac{n}{2^{dm(n)}}\right\}^{1/2} \big(\widehat{H}_n(\mathbf{x}, \varphi) - \overline{H}_n(\mathbf{x}, \varphi)\big) \to N\big(0, \Sigma_\varphi^{*2}(\mathbf{x})\big),$$

*where*

$$\Sigma_\varphi^{2*}(\mathbf{x}) = m_2(\varphi, \mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) \int_{\mathbb{R}^d} \left\{\sum_{k\in\mathbb{Z}^d} \phi(\boldsymbol{k})\phi(\boldsymbol{t}+\boldsymbol{k})\right\}^2 d\mathbf{t}$$

*with* $m_2(\varphi, \mathbf{x})$ *defined in* (4.7).

The proof of Theorem 4.5 is given in Section 6.

**Corollary 4.6.** *Assume that the conditions of Theorem* 4.5 *are satisfied. In addition,* (C.3), (N.4) *and* (N.5) *are assumed to be satisfied. Suppose that*

$$n2^{-(d+2)m(n)} \to 0 \quad as \quad n \to \infty. \tag{4.8}$$

*We have the following convergence in distribution, for* $\mathbf{x} \in D$*, as* $n \to \infty$*,*

$$\left\{\frac{n}{2^{dm(n)}}\right\}^{1/2} \big(\widehat{H}_n(\mathbf{x}) - H(\mathbf{x})\big) \to N\big(0, \Sigma_\varphi^{*2}(\mathbf{x})\big).$$

The proof of Corollary 4.6 is given in Section 6.

The following theorem is more or less straightforward, given Theorem 4.5.

**Theorem 4.7.** *Assume that Condition* (C.2) *is fulfilled. Suppose that*

$$m(n) \to \infty, \quad n2^{-d(m(n))} \to \infty \quad as \;\; n \to \infty.$$

*We have the following convergence in distribution, for* $\mathbf{x} \in D$*, as* $n \to \infty$*,*

$$\left\{\frac{n}{2^{dm(n)}}\right\}^{1/2} \big(\widehat{f}_n(\mathbf{x}) - \overline{f}_n(\mathbf{x})\big) \to N\big(0, \Sigma_\varphi^{**2}(\mathbf{x})\big),$$

*where*

$$\Sigma^{**2}(\mathbf{x}) = f_{\mathbf{X}}(\mathbf{x}) \int_{\mathbb{R}^d} \left\{\sum_{k\in\mathbb{Z}^d} \phi(\boldsymbol{k})\phi(\boldsymbol{t}+\boldsymbol{k})\right\}^2 d\mathbf{t}.$$

**Corollary 4.8.** *Assume that Conditions* (C.1), (C.2), (C.3) *are fulfilled. Suppose that*

$$n2^{-(d+2)m(n)} \to 0 \quad as \ n \to \infty. \tag{4.9}$$

*We have the following convergence in distribution, for* $\mathbf{x} \in D$, *as* $n \to \infty$,

$$\left\{\frac{n}{2^{dm(n)}}\right\}^{1/2} \big(\widehat{f}_n(\mathbf{x}) - f(\mathbf{x})\big) \to N\big(0, \Sigma^{**2}(\mathbf{x})\big).$$

The proof of Corollary 4.8 is given in Section 6.

## 5. CONCLUDING REMARKS AND FUTURE WORK

In this work we have considered the problem of estimation of the density and regression functions in the framework of stationary and ergodic continuous time processes. The uniform strong consistency with the exact rates and the asymptotic normality are obtained by using the sophisticated martingale approach. It is obvious that in some applications the stationarity assumption may be violated, where an important question arises, that is how to extend our results to the setting of non-stationary continuous time processes. The proof of such results, however, should require a different methodology from the one used in the present paper, which goes well beyond the scope of the present paper and leaves this study open for future research.

### *5.1. The Estimation in the Besov Spaces*

According to the Appendix of Masry (2000), there are many equivalent definitions of the Besov spaces $\mathbf{B}_{s,p,q}$, for $s > 0$, $1 \leq p \leq \infty$, and $1 \leq q \leq \infty$. Let

$$(S_{\boldsymbol{\tau}} f)(\mathbf{x}) = f(\mathbf{x} - \boldsymbol{\tau}).$$

For $0 < s < 1$, set

$$\gamma_{s,p,q}(f) = \left(\int_{\mathbb{R}^d}\left(\frac{\|S_{\boldsymbol{\tau}} f - f\|_{L_p}}{\|\boldsymbol{\tau}\|^s}\right)^q \frac{d\boldsymbol{\tau}}{\|\boldsymbol{\tau}\|^d}\right), \qquad \gamma_{s,p,\infty}(f) = \sup_{\boldsymbol{\tau}\in\mathbb{R}^d} \frac{\|S_{\boldsymbol{\tau}} f - f\|_{L_p}}{\|\boldsymbol{\tau}\|^s}.$$

For $s = 1$, set

$$\gamma_{s,p,q}(f) = \left(\int_{\mathbb{R}^d}\left(\frac{\|S_{\boldsymbol{\tau}} f - f\|_{L_p}}{\|\boldsymbol{\tau}\|}\right)^q \frac{d\boldsymbol{\tau}}{\|\boldsymbol{\tau}\|^d}\right), \qquad \gamma_{s,p,\infty}(f) = \sup_{\boldsymbol{\tau}\in\mathbb{R}^d} \frac{\|S_{\boldsymbol{\tau}} f - f\|_{L_p}}{\|\boldsymbol{\tau}\|}.$$

For $0 < s < 1$ and $1 \leq p, q \leq \infty$, define

$$\mathbf{B}_{s,p,q} = \{f \in L_p \colon \gamma_{s,p,q}(f) < \infty\}.$$

For $s > 1$, put

$$s = [s]^- + \{s\}^+$$

with $[s]^-$ an integer and $0 < \{s\}^+ \leq 1$. Define $\mathbf{B}_{s,p,q}$ to be the space of functions in $L_p(\mathbb{R}^d)$ such that $D^j f \in \mathbf{B}_{\{s\}^-,p,q}$ for all $|j| \leq [s]^-$. The norm is defined by

$$\|f\|_{\mathbf{B}_{s,p,q}} = \|f\|_{L_p} + \sum_{|j|\leq[s]^-} \gamma_{\{s\}^-,p,q}(D^j f).$$

For further details, refer to Bergh and Löfström (1967), Triebel (1983) and the Appendix of Masry (2000). Recall that the function $f \in \mathbf{B}_{s,p,q}$ must be in $L_p(\mathbb{R}^d)$ and $s > 0$ is a real-valued smoothness parameter of $f$. A second and very useful characterization of $\mathbf{B}_{s,p,q}$ in terms of wavelets coefficients is due to Meyer (1992). Assume the multiresolution analysis is $r$-regular and $s < r$. Then $f \in \mathbf{B}_{s,p,q}$ if and only if

$$J_{s,p,q}(f) = \|P_{V_0} f\|_{L_p} + \left(\sum_{j>0}\big(2^{js}\|P_{W_j} f\|_{L_p}\big)^q\right)^{1/q} < \infty,$$

with the usual sup-norm modification for $q = \infty$. Moreover, using the wavelet representation, $f \in \mathbf{B}_{s,p,q}$ if and only if

$$J'_{s,p,q}(f) = \|a_0 \cdot\|_{L_p} + \left(\sum_{j>0} \left(2^{j(s+d(1/2-1/p))}\|b_j \cdot\|_{L_p}\right)^q\right)^{1/q} < \infty,$$

where

$$\|b_j \cdot\|_{L_p} = \left(\sum_{i=1}^{d}\sum_{\mathbf{k}\in\mathbb{Z}^d} |b_{i,j,\mathbf{k}}|^p\right)^{1/p}.$$

If we assume that $f, H \in \mathbf{B}_{s,p,q}$, and some additional conditions, one can show, by using similar arguments to those used in Masry (2000), adapted to our setting, that

$$\sup_{\mathbf{x}\in D} |m_n(\mathbf{x},\varphi) - m(\mathbf{x},\varphi)| = O\left(\left(\frac{\log n}{n}\right)^{(s-d/p)/d+2(s-d/p)}\right).$$

This will be considered elsewhere.

## 6. PROOFS

This section is devoted to the proofs of our results. The previously presented notation continues to be used in the following. The following technical lemma will be instrumental in the proof of our theorems.

**Lemma 6.1.** *Let $(Z_n)_{n\geq 1}$ be a sequence of real martingale differences with respect to the sequence of $\sigma$-fields $(\mathcal{F}_n = \sigma(Z_1, \ldots, Z_n))_{n\geq 1}$, where $\sigma(Z_1, \ldots, Z_n)$ is the $\sigma$-field generated by the random variables $Z_1, \ldots, Z_n$. Set*

$$S_n = \sum_{i=1}^{n} Z_i.$$

*For any $p \geq 2$ and any $n \geq 1$, assume that there exist nonnegative constants $C$ and $d_n$ such that*

$$\mathbb{E}\left[Z_n^p \mid \mathcal{F}_{n-1}\right] \leq C^{p-1} p!\, d_n^2, \quad \textit{almost surely.}$$

*Then, for any $\epsilon > 0$, we have*

$$\mathbb{P}(|S_n| > \epsilon) \leq 2\exp\left\{-\frac{\epsilon^2}{2(D_n + C\epsilon)}\right\},$$

*where $D_n = \sum_{i=1}^{n} d_i^2$.*

The proof follows as a particular case of Theorem 8.2.2 due to de la Peña and Giné (1999).

### *Proof of Theorem 3.1*

Define the kernel $K(\mathbf{u},\mathbf{v})$ by

$$K(\mathbf{u},\mathbf{v}) := \sum_{\mathbf{k}\in\mathbb{Z}^d} \phi(\mathbf{u}-\mathbf{k})\phi(\mathbf{v}-\mathbf{k}). \tag{6.1}$$

By using the fact that

$$|\phi(\mathbf{x})| \leq \frac{A_{d+1}}{(1+\|\mathbf{x}\|)^{d+1}},$$

we infer that the kernel function $K(\cdot)$ defined in (6.1) converges uniformly and satisfies Meyer (1992), p. 33,

$$|K(\mathbf{v},\mathbf{u})| \leq \frac{C_{d+1}}{(1+\|\mathbf{v}-\mathbf{u}\|)^{d+1}}, \tag{6.2}$$

for a constant $C_{d+1}$. From (6.2) it follows, for any $j \geq 1$, that

$$\int_{\mathbb{R}^d} |K(\mathbf{v},\mathbf{u})|^j \, d\mathbf{v} \leq G_j(d),$$

where

$$G_j(d) = 2\pi^{d/2} \frac{\Gamma(d)\Gamma(j + d(j-1))}{\Gamma(d/2)\Gamma((d+1)j)} C_{d+1}^j$$

and $\Gamma(t)$ is the Gamma function, that is, $\Gamma(t) := \int_0^\infty y^{t-1} \exp(-y)\, dy$. By combining (3.1), (3.2) and (6.1), we observe that $\widehat{f}_T(\mathbf{x})$ can be written as an extended kernel estimator in the following way

$$\widehat{f}_T(\mathbf{x}) = \frac{1}{Th_T^d} \int_0^T K\left(\frac{\mathbf{x}}{h_T}, \frac{\mathbf{X}_t}{h_T}\right) dt, \qquad \text{where} \quad h_T = 2^{-m(T)}. \tag{6.3}$$

Making use of the triangle inequality, we readily obtain that

$$\begin{aligned}\sup_{\mathbf{x}\in D} \big|\widehat{f}_T(\mathbf{x}) - \mathbb{E}(\widehat{f}_T(\mathbf{x}))\big| &\leq \sup_{\mathbf{x}\in D} \big|\widehat{f}_T(\mathbf{x}) - \overline{f}_T(\mathbf{x})\big| + \sup_{\mathbf{x}\in D} \big|\overline{f}_T(\mathbf{x}) - \mathbb{E}(\widehat{f}_T(\mathbf{x}))\big| \\ &= F_{T,1}(\mathbf{x}) + F_{T,2}(\mathbf{x}), \end{aligned} \tag{6.4}$$

where

$$\overline{f}_T(\mathbf{x}) = \frac{1}{Th_T^d} \int_0^1 \mathbb{E}\left(K\left(\frac{\mathbf{x}}{h_T}, \frac{\mathbf{X}_t}{h_T}\right) \Big| \mathcal{F}_{t-\delta}\right) dt.$$

For a positive real number $\delta$ such that $n = T/\delta \in \mathbb{N}$, consider the $\delta$-partition $\gamma_i = i\delta$, $0 \leq i \leq n$, of the interval $[0,T]$. Moreover, for $t > 0$ and $1 \leq j \leq n$, consider the $\sigma$-fields

$$\mathcal{F}_t = \sigma\big((X_s)\colon 0 \leq s < t\big), \qquad \mathcal{G}_i = \sigma\big((X_s)\colon 0 \leq s \leq T_i\big).$$

Since $D$ is compact, it can be covered by a finite number $L = L(T)$ of cubes $D_j$, with centers $\mathbf{x}_j$ having sides of length

$$r(T) = \text{const}\,/L(T)^{1/d} \qquad \text{for} \quad j = 1,\dots,L(T).$$

We set

$$L(T) = \left(\frac{T}{h_T^{d+2}\log T}\right)^{d/2}.$$

Then we readily infer that

$$\begin{aligned}\sup_{\mathbf{x}\in D} |\widehat{f}_T(\mathbf{x}) - \overline{f}_T(\mathbf{x})| &= \sup_{\mathbf{x}\in D} \left|\frac{1}{Th_T^d} \sum_{i=1}^n \int_{\gamma_{i-1}}^{\gamma_i} \left(K\left(\frac{\mathbf{x}}{h_T}, \frac{\mathbf{X}_t}{h_T}\right) - \mathbb{E}\left[K\left(\frac{\mathbf{x}}{h_T}, \frac{\mathbf{X}_t}{h_T}\right) \Big| \mathcal{F}_{t-\delta}\right]\right) dt\right| \\ &= \max_{1\leq j\leq L(T)} \sup_{\mathbf{x}\in D_j} \left|\frac{1}{Th_T^d} \sum_{i=1}^n \int_{\gamma_{i-1}}^{\gamma_i} \left(K\left(\frac{\mathbf{x}}{h_T}, \frac{\mathbf{X}_t}{h_T}\right) - \mathbb{E}\left[K\left(\frac{\mathbf{x}}{h_T}, \frac{\mathbf{X}_t}{h_T}\right) \Big| \mathcal{F}_{t-\delta}\right]\right) dt\right| \\ &\leq \max_{1\leq j\leq L(T)} \sup_{\mathbf{x}\in D_j} \left|\frac{1}{Th_T^d}\big(Y_i(\mathbf{x}) - Y_i(\mathbf{x}_j)\big)\right| + \max_{1\leq j\leq L(T)} \left|\frac{1}{Th_T^d} Y_i(\mathbf{x}_j)\right| \\ &= I_{T,1}(\mathbf{x}) + I_{T,2}(\mathbf{x}) \end{aligned}$$

with

$$Y_i(\mathbf{x}) := \int_{\gamma_{i-1}}^{\gamma_i} \left(K\left(\frac{\mathbf{x}}{h_T}, \frac{\mathbf{X}_t}{h_T}\right) - \mathbb{E}\left[K\left(\frac{\mathbf{x}}{h_T}, \frac{\mathbf{X}_t}{h_T}\right) \Big| \mathcal{F}_{t-\delta}\right]\right) dt.$$

Since $\phi(\cdot)$ satisfies (2.1) for $|\boldsymbol{\beta}| = 1$, it follows that (Meyer (1992), p. 33)

$$\left|\frac{\partial K(\mathbf{u},\mathbf{y})}{\partial u_i}\right| \leq \frac{C_2}{(1 + \|\mathbf{u}-\mathbf{y}\|)^2} \leq C_2, \quad i = 1,\dots,d.$$

This, in turn, by Cauchy–Schwarz, implies that

$$\left|K(\mathbf{u},\mathbf{y})-K(\mathbf{v},\mathbf{y})\right| \le C_2\sum_{i=1}^{d}|u_i-v_i| \le d^{1/2}C_2\|\mathbf{u}-\mathbf{v}\|.$$

The last equation allows us to infer that

$$\left|K\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{X}_t}{h_T}\Big)-K\Big(\frac{\mathbf{x_j}}{h_T},\frac{\mathbf{X}_t}{h_T}\Big)\right| \le d^{1/2}r(T)C_2h_T^{-1},$$

which implies that

$$I_{T,1}(\mathbf{x}) \le \frac{2d^{1/2}r(T)C_2}{h_T^{d+1}} = \frac{2d^{1/2}\,\mathrm{const}\cdot C_2}{h_T^{d+1}L(T)^{1/d}} = o\Big(\frac{\log T}{Th_T^d}\Big)^{1/2}. \tag{6.5}$$

The proof needs to use Lemma 6.1. Therefore we have to check its conditions. Notice, for any $\delta>0$, that $(Y_i)_{1\le j\le n}$ is a sequence of martingale differences with respect to the sequence of $\sigma$-fields $(\mathcal{G}_{i-1})_{1\le j\le n}$. Indeed, since

$$\mathcal{G}_{i-2}\subseteq\mathcal{F}_{t-\delta}\subseteq\mathcal{G}_{i-1} \qquad \text{for any} \quad t\in[T_{i-1},T_i],$$

it is clear that $Y_i$ is $\mathcal{G}_{i-1}$-measurable and satisfies

$$E\big[Y_i(\mathbf{x})\mid\mathcal{G}_{i-2}\big] = E\bigg[\int_{T_{i-1}}^{T_i}\Big(K\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{X}_t}{h_T}\Big)-\mathbb{E}\Big[K\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{X}_t}{h_T}\Big)\Big|\,\mathcal{F}_{t-\delta}\Big]\Big)\,dt\,\Big|\,\mathcal{G}_{i-2}\bigg]=0.$$

By Minkowski's and Jensen's inequalities, we observe that

$$\begin{aligned}
\big|\mathbb{E}[Y_i^p(\mathbf{x})\mid\mathcal{G}_{i-2}]\big| &\le \mathbb{E}\bigg[\bigg|\int_{\gamma_{i-1}}^{\gamma_i}\Big(K\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{X}_t}{h_T}\Big)-\mathbb{E}\Big[K\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{X}_t}{h_T}\Big)\Big|\,\mathcal{F}_{t-\delta}\Big]\Big)\,dt\bigg|^p\,\bigg|\,\mathcal{G}_{i-2}\bigg]\\
&\le \int_{\gamma_{i-1}}^{\gamma_i}\mathbb{E}\bigg[\Big|K\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{X}_t}{h_T}\Big)-\mathbb{E}\Big[K\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{X}_t}{h_T}\Big)\,\Big|\,\mathcal{F}_{t-\delta}\Big]\Big|^p\,\Big|\,\mathcal{G}_{i-2}\bigg]\,dt\\
&\le \int_{\gamma_{i-1}}^{\gamma_i}\bigg(\mathbb{E}\Big[K^p\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{X}_t}{h_T}\Big)\,\Big|\,\mathcal{G}_{i-2}\Big]^{1/p}+\mathbb{E}\Big[\mathbb{E}\Big[K\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{X}_t}{h_T}\Big)\,\Big|\,\mathcal{F}_{t-\delta}\Big]^p\,\Big|\,\mathcal{G}_{i-2}\Big]^{1/p}\bigg)^p\,dt\\
&\le \int_{\gamma_{i-1}}^{\gamma_i}\bigg(\mathbb{E}\Big[K^p\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{X}_t}{h_T}\Big)\,\Big|\,\mathcal{G}_{i-2}\Big]^{1/p}+\mathbb{E}\Big[\mathbb{E}\Big[K^p\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{X}_t}{h_T}\Big)\,\Big|\,\mathcal{F}_{t-\delta}\Big]\,\Big|\,\mathcal{G}_{i-2}\Big]^{1/p}\bigg)^p\,dt\\
&= \int_{\gamma_{i-1}}^{\gamma_i}\bigg(2\mathbb{E}\Big[K^p\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{X}_t}{h_T}\Big)\,\Big|\,\mathcal{G}_{i-2}\Big]^{1/p}\bigg)^p\,dt\\
&= 2^p\int_{\gamma_{i-1}}^{\gamma_i}\mathbb{E}\Big[K^p\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{X}_t}{h_T}\Big)\,\Big|\,\mathcal{G}_{i-2}\Big]\,dt.
\end{aligned}$$

Observe that

$$\begin{aligned}
E\Big[K^p\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{X}_t}{h_T}\Big)\,\Big|\,\mathcal{G}_{i-2}\Big] &= \int K^p\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{X}_t}{h_T}\Big)f^{\mathcal{G}_{i-2}}(u)\,du\\
&\le \int\frac{C_{d+1}^p}{(1+h_T^{-1}\|x-u\|)^{(d+1)p}}f^{\mathcal{G}_{i-2}}(u)\,du\\
&\le C_{d+1}^p h_T^{(d+1)p}.
\end{aligned} \tag{6.6}$$

Therefore we have

$$\big|E\big[Y_i^p(\mathbf{x})|\mathcal{G}_{i-2}\big]\big| \le 2^p\int_{\gamma_{i-1}}^{\gamma_i}C_{d+1}^p h_T^{(d+1)p}\,dt \le 2\delta C_{d+1}^p h_T^{(d+1)p} \le 2C_{d+1}^{p-1}p!d_i^2,$$

where

$$d_i^2 = \delta C_{d+1}h_T^{(d+1)p}.$$

It follows that

$$D_n = \sum_{j=1}^{n} d_i^2 = Th_T^{(d+1)p} C_{d+1} \le Th_T^d C_{d+1}. \tag{6.7}$$

Therefore we have the following chain of inequalities, for a positive constant $C_2$,

$$\begin{aligned}
&\mathbb{P}\bigg(\max_{1\le j\le L(T)} \bigg|\frac{1}{Th_T^d}\sum_{i=1}^{n} Y_i(\mathbf{x}_j)\bigg| > \epsilon_T\bigg) \le \sum_{j=1}^{L(T)} \mathbb{P}\bigg(\bigg|\sum_{i=1}^{n} Y_n(\mathbf{x}_j)\bigg| > \epsilon_T Th_T^d\bigg)\\
&\le 2L(T)\exp\bigg\{-\frac{(Th_T^d)^2(\log T/Th_T^d)}{2(D_n + 2C_{d+1}Th_T^d(\log T/Th_T^d)^{1/2})}\bigg\}\\
&\le \bigg(\frac{T}{h_T^{d+1}\log T}\bigg)^{d/2}\exp\bigg\{-\frac{Th_T^d\log T}{O(Th_T^d)(1+2C_{d+1}(\log T/Th_T^d)^{1/2})}\bigg\}\\
&\le \bigg(\frac{T}{h_T^{d+1}\log T}\bigg)^{d/2}(T^{-C_2\epsilon_0^2}) = \frac{T^{d/2-C_2\epsilon_0^2}}{(h_T^{d+1}\log T)^{d/2}}\\
&= \frac{1}{\big((Th_T^d)\log T\; h_T T^{2(\epsilon_0^2 C_2/d-2)}\big)^{d/2}}.
\end{aligned}$$

By choosing $\epsilon_0$ sufficiently large, such that

$$\epsilon_0^2 C_2/d - 2 > 0,$$

we readily obtain that

$$\sum_{n\ge 1}\mathbb{P}\bigg(\max_{1\le j\le L(T)}\bigg|\frac{1}{Th_T^d}\sum_{i=1}^{n} Y_i(\mathbf{x}_j)\bigg| > \epsilon_T\bigg) < \infty.$$

We obtain the assertion by a routine application of the Borel–Cantelli lemma

$$I_{T,2}(\mathbf{x}) = O\bigg(\frac{\log T}{Th_T^d}\bigg)^{1/2}. \tag{6.8}$$

We next evaluate the second term in the right-hand side of (6.4). One can see that

$$\begin{aligned}
&\sup_{\mathbf{x}\in D}\big|\overline{f}_T(\mathbf{x}) - \mathbb{E}(\widehat{f}_T(\mathbf{x}))\big|\\
&= \sup_{\mathbf{x}\in D}\bigg|\frac{1}{Th_T^d}\int_0^T \mathbb{E}\Big[K\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{X}_t}{h_T}\Big)\,\Big|\,\mathcal{F}_{t-\delta}\Big]\,dt - \frac{1}{Th_T^d}\int_0^T \mathbb{E}\Big[K\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{X}_t}{h_T}\Big)\Big]\,dt\bigg|\\
&= \sup_{\mathbf{x}\in D}\bigg|\frac{1}{Th_T^d}\int_0^T\int_{\mathbb{R}^d} K\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{y}}{h_T}\Big)\big(f^{\mathcal{F}_{t-\delta}}(\mathbf{y}) - f(\mathbf{y})\big)\,d\mathbf{y}\,dt\bigg|\\
&= \sup_{\mathbf{x}\in D}\bigg|\frac{1}{h_T^d}\int_{\mathbb{R}^d} K\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{y}}{h_T}\Big)\bigg(\frac{1}{T}\int_0^T f^{\mathcal{F}_{t-\delta}}(\mathbf{y})\,dt - f(\mathbf{y})\bigg)\,d\mathbf{y}\bigg|.
\end{aligned}$$

By using the fact that

$$|\phi(\mathbf{x})| \le \frac{A_{d+1}}{(1+\|\mathbf{x}\|)^{d+1}},$$

we infer that the kernel function $K(\cdot)$ defined in (6.1) converges uniformly and satisfies Meyer (1992), p. 33,

$$|K(\mathbf{v},\mathbf{u})| \le \frac{C_{d+1}}{(1+\|\mathbf{v}-\mathbf{u}\|)^{d+1}},$$

for some constant $C_{d+1}$.

Making use of the Cauchy–Schwarz inequality, (6.2) and Condition (C.1), we obtain readily that

$$\begin{aligned}
&\left|\frac{1}{h_T^d}\int_{\mathbb{R}^d} K\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{y}}{h_T}\Big)\Big(\frac{1}{n}\int_0^T f^{\mathcal{F}_{t-\delta}}(\mathbf{y})dt - f(\mathbf{y})\Big)\,d\mathbf{y}\right|\\
&\quad\le \bigg(\int_{\mathbb{R}^d}\left|\frac{1}{h_T^d}K\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{y}}{h_T}\Big)\right|^2 d\mathbf{y}\bigg)^{1/2}\bigg(\int_{\mathbb{R}^d}\left|\frac{1}{T}\int_0^T f^{\mathcal{F}_{t-\delta}}(\mathbf{y})\,dt - f(\mathbf{y})\right|^2 d\mathbf{y}\bigg)^{1/2}\\
&\quad\le \bigg(\int_{\mathbb{R}^d}\Big(\frac{1}{h_T^d}\frac{C_{d+1}}{(1+h_T^{-1}\|\mathbf{x}-\mathbf{y}\|)^{d+1}}\Big)\left|\frac{1}{h_T^d}K\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{y}}{h_T}\Big)\right| d\mathbf{y}\bigg)^{1/2}\left\|\frac{1}{T}\int_0^1 f^{\mathcal{F}_{t-\delta}}\,dt - f\right\|_{L^2}\\
&\quad\le h_T^{1/2}C_{d+1}^{1/2}\int_{\mathbb{R}^d}K\Big(\frac{\mathbf{x}}{h_T},\mathbf{u}\Big)\,du\left\|\frac{1}{T}\int_0^1 f^{\mathcal{F}_{t-\delta}}dt - f\right\|_{L^2}.
\end{aligned}$$

From (6.2) it follows, for any $j \ge 1$, that

$$\int_{\mathbb{R}^d}|K(\mathbf{v},\mathbf{u})|^j\,d\mathbf{v} \le G_j(d),$$

where

$$G_j(d) = 2\pi^{d/2}\frac{\Gamma(d)\Gamma(j+d(j-1))}{\Gamma(d/2)\Gamma((d+1)j)}C_{d+1}^j$$

and $\Gamma(t)$ is the Gamma function, which yields

$$F_{T,2}(\mathbf{x}) \le h_T^{1/2}C_{d+1}^{1/2}G_1^{1/2}(d)\left\|\frac{1}{T}\int_0^1 f^{\mathcal{F}_{t-\delta}}\,dt - f\right\|_{L^2} = O(h_T^{1/2}). \tag{6.9}$$

□

*Proof of Theorem 3.2*

Making use of the triangle inequality, we readily obtain that

$$\begin{aligned}
\sup_{\mathbf{x}\in D}\big|\widehat{f}_T(\mathbf{x}) - f(\mathbf{x})\big| &\le \sup_{\mathbf{x}\in D}\big|\widehat{f}_T(\mathbf{x}) - \mathbb{E}(\widehat{f}_T(\mathbf{x}))\big| + \sup_{\mathbf{x}\in D}\big|\mathbb{E}\big(\widehat{f}_T(\mathbf{x})\big) - f(\mathbf{x})\big|\\
&= F_{T,1}(\mathbf{x}) + F_{T,2}(\mathbf{x}).
\end{aligned} \tag{6.10}$$

By an application of Theorem 3.1 we readily obtain

$$F_{T,1}(\mathbf{x}) = O\bigg(\Big(\frac{\log T}{Th_T^d}\Big)^{1/2}\bigg) + O(h_T^{1/2}). \tag{6.11}$$

Notice that under the conditions imposed in Section 2, we have (refer, e.g., Walter (1994) and Xue (2004)),

$$K(\mathbf{u},\mathbf{v}) = 0 \qquad \text{for} \quad |u_i - v_i| \ge 2L, \quad i = 1,\dots,d,$$

and

$$\int_{\mathbb{R}^d}\bigg\{\prod_{i=1}^d (u_i - v_i)^{k_i}\bigg\}K(\mathbf{u},\mathbf{v})d\mathbf{u} = 0 \qquad \text{for} \quad k_1,\dots,k_d \ge 0, \quad 0 < k_1 + \dots + k_d < r.$$

Under Condition (C.2) and using Taylor series expansion of order $r$ and a change of variables in connection with a straightforward application of Lebesgue dominated convergence theorem, for $\mathbf{x} \in D$ and $0 < \theta < 1$, we have

$$|\mathbb{E}\widehat{f}_T(\mathbf{x}) - f(\mathbf{x})| = \left|\frac{1}{h_T^d}\int_{\mathbb{R}^d}K\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{v}}{h_T}\Big)f(\mathbf{v})d\mathbf{v} - f(\mathbf{x})\right|$$

$$= \Big| \int_{[-2L,2L]^d} K\Big(\frac{\mathbf{x}}{h_T}, \frac{\mathbf{x}}{h_T} + \mathbf{v}\Big) \big(f(\mathbf{x} + h_T\mathbf{v}) - f(\mathbf{x})\big)\, d\mathbf{v} \Big|$$

$$= \Big| \int_{[-2L,2L]^d} K\Big(\frac{\mathbf{x}}{h_T}, \frac{\mathbf{x}}{h_T} + \mathbf{v}\Big) \Big( \frac{1}{r!} \sum_{k_1+\dots+k_d=r} h_T^{k_1} v_1^{k_1} \dots h_T^{k_d} v_d^{k_d} \frac{\partial^r f(\mathbf{v} h_T \theta + \mathbf{x})}{\partial v_1^{k_1} \dots \partial v_d^{k_d}} \Big)\, d\mathbf{v} \Big|$$

$$\le \frac{1}{r!} \sup_{\mathbf{x}\in D} \Big| \frac{\partial^r f(\mathbf{x})}{\partial v_1^{k_1} \dots \partial v_d^{k_d}} \Big| \sum_{k_1+\dots+k_d=r} h_T^{k_1} \dots h_T^{k_d} \int_{[-2L,2L]^d} |v_1^{k_1} \dots v_d^{k_d}| K\Big(\frac{\mathbf{x}}{h_T}, \frac{\mathbf{x}}{h_T} + \mathbf{v}\Big)\, d\mathbf{v}$$

$$= \frac{1}{r!} \sup_{\mathbf{x}\in D} \Big| \frac{\partial^r f(\mathbf{x})}{\partial v_1^{k_1} \dots \partial v_d^{k_d}} \Big| h_T^r \sum_{k_1+\dots+k_d=r} \int_{[-2L,2L]^d} |v_1^{k_1} \dots v_d^{k_d}| \Big| K\Big(\frac{\mathbf{x}}{h_T}, \frac{\mathbf{x}}{h_T} + \mathbf{v}\Big) \Big|\, d\mathbf{v} = O(h_T^r).$$

Therefore

$$F_{T,2}(\mathbf{x}) = \sup_{\mathbf{x}\in D} \big| \mathbb{E}(\widehat{f}_T(\mathbf{x})) - f(\mathbf{x}) \big| = O(h_T^r). \tag{6.12}$$

By combining (6.11) and (6.12) we achieve the proof of Theorem 3.2. □

*Proof of Theorem 4.1*

Let us introduce the truncated version of $\widehat{H}_T(\mathbf{x})$ as follows. Let

$$\widehat{a}_{m\mathbf{k}}^L = \frac{1}{T} \int_0^T \varphi(Y_t) \mathbf{1}\{|\varphi(Y_t)| \le L_T\} \phi_{m,\mathbf{k}}(\mathbf{X}_t)\, dt \tag{6.13}$$

and

$$\widehat{H}_T^L(\mathbf{x}) = \sum_{\mathbf{k}\in\mathbb{Z}^d} \widehat{a}_{m\mathbf{k}}^L \phi_{m,\mathbf{k}}(\mathbf{x}). \tag{6.14}$$

Here and in the sequel, $\mathbf{1}\{A\}$ denotes the indicator function of the set $A$. In a similar way as in the preceding proof, we write $\widehat{H}_T(\mathbf{x})$ as an extended kernel estimator

$$\widehat{H}_T(\mathbf{x}) = \frac{1}{T h_T^d} \int_0^T \varphi(Y_t) K\Big(\frac{\mathbf{x}}{h_T}, \frac{\mathbf{X}_t}{h_T}\Big)\, dt, \tag{6.15}$$

and

$$\widehat{H}_T^L(\mathbf{x}) = \frac{1}{T h_T^d} \int_0^T \varphi(Y_t) \mathbf{1}\{|\varphi(Y_t)| \le L_T\} K\Big(\frac{\mathbf{x}}{h_T}, \frac{\mathbf{X}_t}{h_T}\Big)\, dt. \tag{6.16}$$

We recall

$$\overline{H}_T(\mathbf{x}) = \frac{1}{T h_T^d} \int_0^T \mathbb{E}\Big[ \varphi(Y_t) K\Big(\frac{\mathbf{x}}{h_T}, \frac{\mathbf{X}_t}{h_T}\Big) \Big| \mathcal{G}_{t-\delta} \Big]\, dt,$$

and

$$\overline{H}_T^L(\mathbf{x}) = \frac{1}{T h_T^d} \int_0^T \mathbb{E}\Big[ \varphi(Y_t) \mathbf{1}\{|\varphi(Y_t)| \le L_T\} K\Big(\frac{\mathbf{x}}{h_T}, \frac{\mathbf{X}_t}{h_T}\Big) \Big| \mathcal{G}_{t-\delta} \Big]. \tag{6.17}$$

We first decompose $\big\{\widehat{H}_T(\mathbf{x}) - \mathbb{E}(\widehat{H}_T(\mathbf{x}))\big\}$ into the sum of four components by writing

$$\widehat{H}_T(\mathbf{x}) - \overline{H}_T(\mathbf{x}) = \big(\widehat{H}_T(\mathbf{x}) - \widehat{H}_T^L(\mathbf{x})\big) + \big(\widehat{H}_T^L(\mathbf{x}) - \overline{H}_T^L(\mathbf{x})\big) + \big(\overline{H}_T^L(\mathbf{x}) - \overline{H}_T(\mathbf{x})\big).$$

Recalling (6.2), we observe that

$$\sup_{\mathbf{x}\in D} \big| \widehat{H}_T(\mathbf{x}) - \widehat{H}_T^L(\mathbf{x}) \big| \le \frac{h_T C_{d+1}}{T} \int_0^T \big| \varphi(Y_t) \mathbf{1}\{|\varphi(Y_t)| > L_T\} \big|\, dt. \tag{6.18}$$

A simple Markov inequality application implies that

$$\mathbb{P}\big(|\varphi(Y_t)| > L_T\big) \le L_T^{-\nu}\mathbb{E}\big(|\varphi(Y_t)|^{\nu}\big).$$

Recall the notation $T = n\delta$, under Assumption (N.0) and using the fact that

$$\sum_{n>1} L_T^{-\nu} < \infty, \qquad T \to \infty.$$

We have by the Borel–Cantelli lemma that $|\varphi(Y_t)| \le L_T$ almost surely for all sufficiently large $T$. Since $L_T$ is increasing,

$$|\varphi(Y_t)| \le L_T \qquad \text{for all} \quad t \le T.$$

By all this and using (6.18) we infer that

$$\sup_{\mathbf{x}\in D}\big|\widehat{H}_T(\mathbf{x}) - \widehat{H}_T^L(\mathbf{x})\big| = o(1) \;\; a.s. \tag{6.19}$$

Once more, by (6.2) we infer that

$$\begin{aligned}
&\big|\overline{H}_T^L(\mathbf{x}) - \overline{H}_T(\mathbf{x})\big| \\
&\quad \le \frac{1}{Th_T^d}\int_0^T \left|\mathbb{E}\left[\varphi(Y_t)\mathbf{1}\{|\varphi(Y_t)| \le L_T\}K\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{X}_t}{h_T}\Big)\;\Big|\;\mathcal{G}_{t-\delta}\right] - \mathbb{E}\left[\varphi(Y_t)K\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{X}_t}{h_T}\Big)\;\Big|\;\mathcal{G}_{t-\delta}\right]\right| dt \\
&\quad \le \frac{1}{Th_T^d}\int_0^T \mathbb{E}\left[|\varphi(Y_t)|\mathbf{1}\{|\varphi(Y_t)| > L_T\}\Big|K\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{X}_t}{h_T}\Big)\Big|\;\Big|\;\mathcal{G}_{t-\delta}\right] dt \\
&\quad \le \frac{1}{Th_T^d}\int_0^T \mathbb{E}\left[|\varphi(Y_t)|\mathbf{1}\{|\varphi(Y_t)| > L_T\}\frac{C_{d+1}}{h_T^d(1+\|\mathbf{x}-\mathbf{X}_t\|h_T^{-1})^{d+1}}\;\Big|\;\mathcal{G}_{t-\delta}\right] dt \\
&\quad \le \frac{h_T C_{d+1}}{T}\int_0^T \mathbb{E}\Big[|\varphi(Y_t)|\mathbf{1}\{|\varphi(Y_t)| > L_T\}\mid \mathcal{G}_{t-\delta}\Big]\, dt.
\end{aligned}$$

By the Hölder and Markov inequalities, we obtain, for any $\epsilon > 0$ and any $p$ and $q$ fulfilling

$$\frac{1}{p} + \frac{1}{q} = 1,$$

that

$$\begin{aligned}
\mathbb{E}\big[|\varphi(Y_t)|\mathbf{1}\{|\varphi(Y_t)| > L_T\}\mid \mathcal{G}_{t-\delta}\big] &\le \big(\mathbb{E}\big[|\varphi(Y_t)|^q\big|\mathcal{G}_{t-\delta}\big]\big)^{1/q}\big(\mathbb{P}\{|\varphi(Y_t)| > L_T\big|\mathcal{G}_{t-\delta}\}\big)^{1/p} \\
&\le L_T^{-q}\mathbb{E}\big[|\varphi(Y_t)|^q \mid \mathcal{G}_{t-\delta}\big] = L_T^{-q}\int_{\mathbb{R}}|\varphi(v)|^q\rho^{\mathcal{G}_{t-\delta}}(v)\,dv.
\end{aligned}$$

The process $(\rho^{\mathcal{G}_{t-\delta}}(v))_{t\in\mathbb{T}}$ fulfills Condition (N.1), which combined with Condition (N.0), in turn, readily implies that

$$\begin{aligned}
\sup_{\mathbf{x}\in D}\big|\overline{H}_T^L(\mathbf{x}) - \overline{H}_T(\mathbf{x})\big| &\le h_T\sup_{y\in\mathbb{R}}\left\|\frac{1}{T\rho(y)}\int_0^T\rho^{\mathcal{G}_{t-\delta}}(y)\right\|L_T^{-q}\int_{\mathbb{R}}|\varphi(v)|^q\rho(v)\,dv \\
&= h_T\sup_{y\in\mathbb{R}}\left\|\frac{1}{T\rho(y)}\int_0^T\rho^{\mathcal{G}_{t-\delta}}(y)\right\|L_T^{-q}\mathbb{E}\big[|\varphi(Y_0)|^q\big],
\end{aligned}$$

which gives

$$\sup_{\mathbf{x}\in D}\big|\overline{H}_T^L(\mathbf{x}) - \overline{H}_t(\mathbf{x})\big| = O(h_T) \;\; a.s. \tag{6.20}$$

We set

$$D(T) = \Big\lfloor\frac{TL_T^2}{h_T^{d+2}\log T}\Big\rfloor^{d/2}.$$

Recalling that the set $D$ is compact and using the same arguments as in the proof of Theorem 3.1, we obtain that

$$\begin{aligned}
\sup_{\mathbf{x}\in D}\left|\widehat{H}_T^L(\mathbf{x})-\overline{H}_T^L(\mathbf{x})\right| &= \max_{1\le j\le D(T)}\sup_{\mathbf{x}\in D\cap I_j}\left|\widehat{H}_T^L(\mathbf{x})-\overline{H}_T^L(\mathbf{x})\right| \\
&\le \max_{1\le j\le D(T)}\sup_{\mathbf{x}\in D\cap I_j}\left|\widehat{H}_T^L(\mathbf{x})-\widehat{H}_n^T(\mathbf{x}_j)\right| + \max_{1\le j\le L(n)}\left|\widehat{H}_n^T(\mathbf{x}_j)-\overline{H}_n^T(\mathbf{x}_j)\right| \\
&\quad + \max_{1\le j\le L(n)}\sup_{\mathbf{x}\in D\cap I_j}\left|\overline{H}_n^T(\mathbf{x}_i)-\overline{H}_n^T(\mathbf{x})\right| \\
&= Q_1+Q_2+Q_3. \qquad (6.21)
\end{aligned}$$

From (6.12), we infer that, almost surely,

$$\begin{aligned}
\left|\widehat{H}_T^L(\mathbf{x})-\widehat{H}_T^L(\mathbf{x}_j)\right| &= \left|\frac{1}{Th_T^d}\int_0^T \varphi(Y_t)\mathbf{1}\{|\varphi(Y_t)|\le L_T\}\Big(K\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{X}_t}{h_T}\Big)-K\Big(\frac{\mathbf{x}_j}{h_T},\frac{\mathbf{X}_t}{h_T}\Big)\Big)\,dt\right| \\
&\le \frac{1}{Th_T^d}\int_0^T|\varphi(Y_t)|\left|K\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{X}_t}{h_T}\Big)-K\Big(\frac{\mathbf{x}_j}{h_T},\frac{\mathbf{X}_t}{h_T}\Big)\right| \\
&= \frac{d^{1/2}C_2L_T}{h_T^{d+1}}\|\mathbf{x}-\mathbf{x}_j\|.
\end{aligned}$$

This, in turn, implies that

$$\begin{aligned}
Q_1 &= \max_{1\le j\le D(T)}\sup_{\mathbf{x}\in D\cap I_j}\left|\widehat{H}_T^L(\mathbf{x})-\widehat{H}_T^L(\mathbf{x}_j)\right| \\
&\le \text{const}\cdot\frac{L_T}{L^{1/d}(T)h_T^{d+1}} = O\Big(\Big(\frac{\log T}{Th_T^d}\Big)^{1/2}\Big)\quad a.s. \qquad (6.22)
\end{aligned}$$

In a similar way, making use of (6.12) gives

$$\begin{aligned}
&\left|\overline{H}_T^L(\mathbf{x})-\overline{H}_T^L(\mathbf{x}_j)\right| \\
&= \left|\frac{1}{Th_T^d}\int_0^T\mathbb{E}\Big[\varphi(Y_t)\mathbf{1}\{|\varphi(Y_t)|\le L_T\}\Big(K\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{X}_t}{h_T}\Big)-K\Big(\frac{\mathbf{x}_j}{h_T},\frac{\mathbf{X}_t}{h_T}\Big)\Big)\,\Big|\,\mathcal{G}_{t-\delta}\Big]\,dt\right| \\
&\le \frac{1}{nh_n^d}\int_0^T\mathbb{E}\Big[|\varphi(Y_t)|\left|K\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{X}_t}{h_T}\Big)-K\Big(\frac{\mathbf{x}_j}{h_T},\frac{\mathbf{X}_t}{h_T}\Big)\right|\,\Big|\,\mathcal{G}_{t-\delta}\Big]\,dt \\
&= \frac{d^{1/2}C_2L_T}{h_T^{d+1}}\|\mathbf{x}-\mathbf{x}_j\|.
\end{aligned}$$

This yields likewise

$$\begin{aligned}
Q_3 &= \max_{1\le j\le D(T)}\sup_{\mathbf{x}\in D\cap I_j}\left|\overline{H}_T^L(\mathbf{x}_i)-\overline{H}_T^L(\mathbf{x})\right| \\
&\le \text{const}\cdot\frac{L_T}{L^{1/d}(T)h_T^{d+1}} = O\Big(\Big(\frac{\log T}{Th_T^d}\Big)^{1/2}\Big)\quad a.s. \qquad (6.23)
\end{aligned}$$

We next evaluate the term $W_2$ in the right-hand side of (6.21). Observe that

$$\begin{aligned}
Q_2 &= \max_{1\le j\le D(T)}\left|\widehat{H}_T^L(\mathbf{x}_j)-\overline{H}_T^L(\mathbf{x}_j)\right| \\
&= \max_{1\le j\le D(T)}\left|\frac{1}{Th_T^d}\int_0^T\Big(\varphi(Y_t)\mathbf{1}\{|\varphi(Y_t)|\le L_T\}K\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{X}_t}{h_T}\Big)\right. \\
&\qquad \left. -\,\mathbb{E}\Big[\varphi(Y_t)\mathbf{1}\{|\varphi(Y_t)|\le L_T\}K\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{X}_t}{h_T}\Big)\,\Big|\,\mathcal{G}_{t-\delta}\Big]\Big)\,dt\right|
\end{aligned}$$

$$= \max_{1\le j\le D(T)} \left| \frac{1}{Th_T^d} \sum_{i=1}^{n} V_{T,i}(\mathbf{x}_j) \right|,$$

where, for $i = 1, \ldots, n$,

$$V_{T,i}(\mathbf{x}) = \int_{T_{i-1}}^{T_i} \varphi(Y_t)\mathbf{1}\{|\varphi(Y_t)| \le L_T\} K\Big(\frac{\mathbf{x}}{h_T}, \frac{\mathbf{X}_t}{h_T}\Big) - \mathbb{E}\Big[\varphi(Y_t)\mathbf{1}\{|\varphi(Y_t)| \le L_T\} K\Big(\frac{\mathbf{x}}{h_T}, \frac{\mathbf{X}_t}{h_T}\Big) \,\Big|\, \mathcal{G}_{t-\delta}\Big]\, dt$$

is a sequence of martingale differences array with respect to the $\sigma$-field $\mathcal{G}_i$. It is easy to see that, using Jensen's and Minkowski's inequalities, for $i = 1, \ldots, n$,

$$\begin{aligned}
\mathbb{E}(V_{T,i}^p(\mathbf{x}) \mid \mathcal{G}_{i-2}) &= \mathbb{E}\bigg(\bigg(\int_{T_{i-1}}^{T_i} \Big(\varphi(Y_t)\mathbf{1}\{|\varphi(Y_t)| \le L_T\} K\Big(\frac{\mathbf{x}}{h_T}, \frac{\mathbf{X}_t}{h_T}\Big) \\
&\quad - \mathbb{E}\Big[\varphi(Y_t)\mathbf{1}\{|\varphi(Y_t)| \le L_T\} K\Big(\frac{\mathbf{x}}{h_T}, \frac{\mathbf{X}_t}{h_T}\Big) \,\Big|\, \mathcal{G}_{t-\delta}\Big]\Big)\, dt\bigg)^p \,\bigg|\, \mathcal{G}_{i-2}\bigg) \\
&\le \int_{T_{i-1}}^{T_i} \bigg(\mathbb{E}^{1/p}\Big(\Big(\varphi(Y_t)\mathbf{1}\{|\varphi(Y_t)| \le L_T\} K\Big(\frac{\mathbf{x}}{h_T}, \frac{\mathbf{X}_t}{h_T}\Big)\Big)^p \,\Big|\, \mathcal{G}_{i-2}\Big) \\
&\quad + \mathbb{E}^{1/p}\Big(\mathbb{E}\Big[\varphi(Y_t)\mathbf{1}\{|\varphi(Y_t)| \le L_T\} K\Big(\frac{\mathbf{x}}{h_T}, \frac{\mathbf{X}_t}{h_T}\Big) \,\Big|\, \mathcal{G}_{t-\delta}\Big]\Big)^p \,\Big|\, \mathcal{G}_{i-2}\bigg)^p dt. \qquad (6.24)
\end{aligned}$$

Recalling that for $\delta > 0$ and $T_{i-1} \le t \le T_i$, $\mathcal{G}_{i-2} \subseteq \mathcal{G}_{t-\delta} \subseteq \mathcal{G}_{i-1}$

$$\begin{aligned}
&\mathbb{E}\Big[\Big(\mathbb{E}\Big[\varphi(Y_t)\mathbf{1}\{|\varphi(Y_t)| \le L_T\} K\Big(\frac{\mathbf{x}}{h_T}, \frac{\mathbf{X}_t}{h_T}\Big) \,\Big|\, \mathcal{G}_{t-\delta}\Big]\Big)^p \,\Big|\, \mathcal{G}_{i-2}\Big] \\
&\quad \le \mathbb{E}\Big[\Big(\varphi(Y_t)\mathbf{1}\{|\varphi(Y_t)| \le L_T\} K\Big(\frac{\mathbf{x}}{h_T}, \frac{\mathbf{X}_t}{h_T}\Big)\Big)^p \,\Big|\, \mathcal{G}_{i-2}\Big]
\end{aligned}$$

is $\mathcal{G}_{i-1}$-measurable and using (6.24), it follows that

$$\mathbb{E}[V_{T,i}^p(\mathbf{x}) \mid \mathcal{G}_{i-2}] \le 2^p \int_{T_{i-1}}^{T_i} \mathbb{E}\Big[\Big(\varphi(Y_t)\mathbf{1}\{|\varphi(Y_t)| \le L_T\} K\Big(\frac{\mathbf{x}}{h_T}, \frac{\mathbf{X}_t}{h_T}\Big)\Big)^p \,\Big|\, \mathcal{G}_{i-2}\Big]\, dt \qquad (6.25)$$

$$\le L_T^p 2^p \int_{T_{i-1}}^{T_i} \mathbb{E}\Big[\Big|K^p\Big(\frac{\mathbf{x}}{h_T}, \frac{\mathbf{X}_t}{h_T}\Big)\Big| \,\Big|\, \mathcal{G}_{i-2}\Big]\, dt. \qquad (6.26)$$

From (6.6) we obtain, for any integer $p$, that

$$\big|\mathbb{E}\big[V_{T,i}^p(\mathbf{x})\big|\mathcal{G}_{i-2}\big]\big| \le \delta L_T^p 2^p C_{d+1}^p h_T^{p(d+1)} \le p! C^{p-1} d_i^2.$$

Taking

$$C = 2C_{d+1}, \qquad d_i^2 = \delta C_{d+1} L_T^p h_T^{p(d+1)},$$

and making use of the condition

$$L_T^p 2^{-m(n)(d(p-1)+p)} = O(1) \quad \text{as } n \to \infty,$$

we conclude that

$$D_n = \sum_{i=1}^{n} d_i^2 = n\delta L_T^p h_T^{p(d+1)} = Th_T^d (L_T^p h_T^{d(p-1)+p}) = O(Th_T^d).$$

Applying Lemma 6.1 to the sum of $\{V_{T,i}(\mathbf{x}_j)\}$ and choosing

$$\epsilon_T = \epsilon_0 (\log T / Th_T^d)^{1/2},$$

we obtain

$$\mathbb{P}\Big(\max_{1\le j\le D(T)}\Big|\frac{1}{Th_T^d}\sum_{i=1}^{n}V_{T,i}(\mathbf{x}_j)\Big|>\epsilon_T\Big)\le\sum_{j=1}^{D(T)}\mathbb{P}\Big(\Big|\sum_{i=1}^{n}V_{T,i}(\mathbf{x}_j)\Big|>\epsilon_T Th_T^d\Big)$$
$$\le 2D(T)\exp\Big\{-\frac{(Th_T^d)^2(\log T/Th_T^d)}{2(D_n+2C_{d+1}Th_T^d(\log T/Th_T^d)^{1/2})}\Big\}$$
$$\le\Big(\frac{TL_T^2}{h_T^{d+1}\log T}\Big)^{d/2}\exp\Big\{-\frac{Th_T^d\log T}{O(Th_T^d)(1+2C_{d+1}(\log T/Th_T^d)^{1/2})}\Big\}$$
$$\le\Big(\frac{TL_T^2}{h_T^{d+1}\log T}\Big)^{d/2}(T^{-C_2\epsilon_0^2}).$$

We conclude that

$$\max_{1\le j\le D(T)}\big|\widehat{H}_T^L(\mathbf{x}_j)-\overline{H}_T^L(\mathbf{x}_j)\big|=O\Big(\Big(\frac{\log T}{Th_T^d}\Big)^{1/2}\Big) \tag{6.27}$$

by a routine application of the Borel–Cantelli lemma. This, when combined with (6.22), (6.27) and (6.23), and the fact that $h_T=2^{-m(T)}$, implies the desired result

$$\sup_{\mathbf{x}\in D}\big|\widehat{H}_T(\mathbf{x})-\overline{H}_T(\mathbf{x})\big|=O\Big(\Big(\frac{(\log T)2^{dm(T)}}{T}\Big)^{1/2}\Big). \tag{6.28}$$

Therefore the proof of Theorem 3.2 is completed by combining (6.19), (6.20) and (6.28). □

*Proof of Theorem 4.2*

Consider the decomposition

$$\widehat{H}_T(\mathbf{x})-H(\mathbf{x})=\big(\widehat{H}_T(\mathbf{x})-\overline{H}_T(\mathbf{x})\big)+\big(\overline{H}_T(\mathbf{x})-\mathbb{E}(\widehat{H}_T(\mathbf{x}))\big)+\big(\mathbb{E}(\widehat{H}_T(\mathbf{x}))-H(\mathbf{x})\big)$$
$$=\mathcal{K}_{T,1}(\mathbf{x})+\mathcal{K}_{T,2}(\mathbf{x})+\mathcal{K}_{T,3}(\mathbf{x}). \tag{6.29}$$

Applying Theorem 3.2, we have

$$\sup_{\mathbf{x}\in D}|\mathcal{K}_{T,1}(\mathbf{x})|=O\Big(\Big(\frac{\log T}{Th_T^d}\Big)^{1/2}\Big)+O(h_T)\ \ a.s. \tag{6.30}$$

Making use of the Cauchy–Schwarz inequality in combination with (6.2), we infer that

$$\big|\overline{H}_T(\mathbf{x})-\mathbb{E}(\widehat{H}_T(\mathbf{x}))\big|\le\frac{1}{Th_T^d}\int_0^T\int_{\mathbb{R}^{d+1}}\varphi(v)K\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{u}}{h_T}\Big)\big(g^{\mathcal{G}_{t-\delta}}(\mathbf{u},v)-g(\mathbf{u},v)\big)\,d\mathbf{u}\,dv\,dt$$
$$\le\Big\|\frac{1}{T}\int_0^T g^{\mathcal{G}_{t-\delta}}\,dt-g\Big\|_{L^2}\Big(\int_{\mathbb{R}^{d+1}}\Big(\frac{C_{d+1}}{h_T^d(1+\|x-u\|h_T^{-1})^{d+1}}\Big)\varphi^2(v)\Big(\frac{1}{h_T^d}K\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{u}}{h_T}\Big)\Big)\,d\mathbf{u}\,dv\Big)^{1/2}$$
$$\le h_T^{1/2}\Big\|\frac{1}{T}\int_0^T g^{\mathcal{G}_{t-\delta}}\,dt-g\Big\|_{L^2}\|\varphi\|_{L^2}\Big(\int_{\mathbb{R}^d}\Big|K\Big(\frac{\mathbf{x}}{h_T},\mathbf{z}\Big)\Big|\,d\mathbf{z}\Big)^{1/2}$$
$$\le h_T^{1/2}C_{(d+1)}^2G_1^{1//2}(d)\|\varphi\|_{L^2}\Big\|\frac{1}{T}\int_0^T g^{\mathcal{G}_{t-\delta}}\,dt-g\Big\|_{L^2}.$$

Under condition (N.3), we obtain

$$\sup_{\mathbf{x}\in D}|\mathcal{K}_{T,2}(\mathbf{x})|=\sup_{\mathbf{x}\in D}\big|\overline{H}_T(\mathbf{x})-\mathbb{E}(\widehat{H}_T(\mathbf{x}))\big|=O(h_T^{1/2})\ \ a.s. \tag{6.31}$$

By change of variables, it follows that

$$\begin{aligned}
\big|\mathbb{E}(\widehat{H}_T(\mathbf{x})) - H(\mathbf{x})\big| &= \frac{1}{Th_T^d}\int_0^T \mathbb{E}\Big[\varphi(Y_t)K\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{X}_t}{h_T}\Big)\Big]\,dt - H(\mathbf{x})\\
&= \frac{1}{Th_T^d}\int_0^T \mathbb{E}\Big[\mathbb{E}[\varphi(Y_t)\mid \mathbf{X}_t]K\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{X}_t}{h_T}\Big)\Big] - H(\mathbf{x})\\
&= \frac{1}{Th_T^d}\int_0^T \mathbb{E}\Big[m(\varphi,\mathbf{X}_t)K\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{X}_t}{h_T}\Big)\Big] - m(\varphi,\mathbf{x})f(\mathbf{x})\\
&= \frac{1}{h_T^d}\int_{\mathbb{R}^d} m(\varphi,\mathbf{u})K\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{u}}{h_T}\Big)f(\mathbf{u})\,d\mathbf{u} - m(\varphi,\mathbf{x})f(\mathbf{x})\\
&= \int_{[-2L,2L]^d} m(\varphi,\mathbf{x}+h_T\mathbf{v})f(\mathbf{x}+h_T\mathbf{v})K\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{x}}{h_T}+\mathbf{v}\Big)\,d\mathbf{v} - m(\varphi,\mathbf{x})f(\mathbf{x})\\
&= \int_{[-2L,2L]^d} \big(m(\varphi,\mathbf{x}+h_T\mathbf{v})f(\mathbf{x}+h_T\mathbf{v}) - m(\varphi,\mathbf{x})f(\mathbf{x})\big)K\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{x}}{h_T}+\mathbf{v}\Big)\,d\mathbf{v}.
\end{aligned}$$

Making use of Conditions (C.2), (N.2) and using the same arguments, we infer that

$$\begin{aligned}
&\big|\mathbb{E}(\widehat{H}_T(\mathbf{x})) - H(\mathbf{x})\big|\\
&\leq \frac{1}{r!}\sup_{\mathbf{x}\in D}\left|\frac{\partial^r m(\varphi,\mathbf{x})f(\mathbf{x})}{\partial v_1^{k_1}\ldots\partial v_d^{k_d}}\right| \sum_{k_1+\cdots+k_d=r} h_T^{k_1}\ldots h_T^{k_d}\int_{[-2L,2L]^d}|v_1^{k_1}\ldots v_d^{k_d}|K\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{x}}{h_T}+\mathbf{v}\Big)\,d\mathbf{v}\\
&= O(h_T^r).
\end{aligned}$$

Therefore

$$\sup_{\mathbf{x}\in D}\big|\mathcal{K}_{T,3}(\mathbf{x})\big| = \sup_{\mathbf{x}\in D}\big|\mathbb{E}(\widehat{H}_T(\mathbf{x})) - H(\mathbf{x})\big| = O(h_T^r). \tag{6.32}$$

The proof of Theorem 4.2 is completed by combining (6.30), (6.31) and (6.32). □

*Proof of Theorem 4.3*

Consider the decomposition

$$\begin{aligned}
&\sup_{\mathbf{x}\in D}\big|m_T(\mathbf{x},\varphi) - m(\mathbf{x},\varphi)\big|\\
&\quad= \sup_{\mathbf{x}\in D}\left|\frac{\widehat{H}_T(\mathbf{x},\varphi)}{\widehat{f}_T(\mathbf{x})} - \frac{H(\mathbf{x},\varphi)}{\widehat{f}_T(\mathbf{x})} + \frac{H(\mathbf{x},\varphi)}{\widehat{f}_T(\mathbf{x})} - \frac{H(\mathbf{x},\varphi)}{f(\mathbf{x})}\right|\\
&\quad\leq \sup_{\mathbf{x}\in D}\left|\frac{\widehat{H}_T(\mathbf{x},\varphi) - H(\mathbf{x},\varphi)}{\widehat{f}_T(\mathbf{x})}\right| + \sup_{\mathbf{x}\in D}\left|\frac{H(\mathbf{x},\varphi)}{f(\mathbf{x})}\;\frac{f(\mathbf{x}) - \widehat{f}_T(\mathbf{x})}{\widehat{f}_T(\mathbf{x})}\right|\\
&\quad\leq \frac{\sup_{\mathbf{x}\in D}|\widehat{H}_T(\mathbf{x},\varphi) - H(\mathbf{x},\varphi)|}{\inf_{\mathbf{x}\in D}|\widehat{f}_T(\mathbf{x})|} + \sup_{\mathbf{x}\in D}|m(\mathbf{x},\varphi)|\frac{\sup_{\mathbf{x}\in D}|\widehat{f}_T(\mathbf{x}) - f(\mathbf{x})|}{\inf_{\mathbf{x}\in D}|\widehat{f}_T(\mathbf{x})|}.
\end{aligned} \tag{6.33}$$

Observe that

$$\inf_{\mathbf{x}\in D}|\widehat{f}_T(\mathbf{x})| \geq \inf_{\mathbf{x}\in D}|f(\mathbf{x})| - \sup_{\mathbf{x}\in D}|\widehat{f}_T(\mathbf{x}) - f(\mathbf{x})|.$$

Theorem 3.2 combined with Condition (N.4)(i) imply that

$$\inf_{\mathbf{x}\in D}|\widehat{f}_T(\mathbf{x})| \geq \inf_{\mathbf{x}\in D}|f(\mathbf{x})| > \lambda > 0. \tag{6.34}$$

Under Condition (N.4)(ii) we have

$$\sup_{\mathbf{x}\in D}|m(\mathbf{x},\varphi)| \le \Lambda < \infty. \tag{6.35}$$

Theorems 3.2 and 4.2 combined with statements (6.34) and (6.35) finish the proof of Theorem 4.3. $\square$

*Proof of Theorem 4.4*

Let us define some notation needed:

$$\widetilde{f}_T(\mathbf{x}) = \frac{1}{Th_T^d}\int_0^T \mathbb{E}\Big[K\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{X}_t}{h_T}\Big)\,\Big|\,\mathcal{G}_{t-\delta}\Big]\,dt,$$

$$\mathcal{Q}_T(\mathbf{x}) = \widehat{H}_T(\mathbf{x}) - \overline{H}_T(\mathbf{x}) + m(\mathbf{x},\varphi)\big(\widehat{f}_T(\mathbf{x}) - \widetilde{f}_T(\mathbf{x})\big), \tag{6.36}$$

$$\mathcal{B}_T(\mathbf{x}) = \overline{m}_T(\mathbf{x},\varphi) - m(\mathbf{x},\varphi) = \frac{\overline{H}_T(\mathbf{x})}{\widetilde{f}_T(\mathbf{x})} - m(\mathbf{x},\varphi), \tag{6.37}$$

$$\mathcal{R}_T(\mathbf{x}) = -\mathcal{B}_T(\mathbf{x})\big(\widehat{f}_T(\mathbf{x}) - \widetilde{f}_T(\mathbf{x}\big). \tag{6.38}$$

We may consider then the decomposition

$$\begin{aligned} m_T(\mathbf{x},\varphi) - m(\mathbf{x},\varphi) &= \big\{m_T(\mathbf{x},\varphi) - \overline{m}_T(\mathbf{x},\varphi)\big\} + \big\{\overline{m}_T(\mathbf{x},\varphi) - m(\mathbf{x},\varphi)\big\} \\ &= \frac{\mathcal{Q}_T(\mathbf{x}) + \mathcal{R}_T(\mathbf{x})}{\widehat{f}_T(\mathbf{x})} + \mathcal{B}_T(\mathbf{x}). \end{aligned} \tag{6.39}$$

By using Condition (N.5), we obtain

$$\begin{aligned} \mathcal{B}_T(\mathbf{x}) &= \overline{m}_T(\mathbf{x},\varphi) - m(\mathbf{x},\varphi) = \frac{\widetilde{H}_T(\mathbf{x})}{\widetilde{f}_T(\mathbf{x})} - m(\mathbf{x},\varphi) \\ &= \frac{1}{\widetilde{f}_T(\mathbf{x})}\Big\{\frac{1}{Th_T^d}\int_0^T \mathbb{E}\Big[\big(\varphi(Y_t) - m(\mathbf{x},\varphi)\big)K\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{X}_t}{h_T}\Big)\,\Big|\,\mathcal{G}_{t-\delta}\Big]\,dt\Big\} \\ &= \frac{1}{\widetilde{f}_T(\mathbf{x})}\Big\{\frac{1}{Th_T^d}\int_0^T \mathbb{E}\Big[\mathbb{E}\Big[\big(\varphi(Y_t) - m(\mathbf{x},\varphi)\big)K\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{X}_t}{h_T}\Big)\,\Big|\,\mathcal{S}_{t-\delta,\delta}\Big]\,\Big|\,\mathcal{G}_{t-\delta}\Big]\,dt\Big\} \\ &= \frac{1}{\widetilde{f}_T(\mathbf{x})}\Big\{\frac{1}{Th_T^d}\int_0^T \mathbb{E}\Big[\big(\mathbb{E}[\varphi(Y_t)\mid \mathbf{X}_t] - m(\mathbf{x},\varphi)\big)K\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{X}_t}{h_T}\Big)\,\Big|\,\mathcal{G}_{t-\delta}\Big]\,dt\Big\} \\ &= \frac{1}{\widetilde{f}_T(\mathbf{x})}\Big\{\frac{1}{Th_T^d}\int_0^T \mathbb{E}\Big[\big(m(\mathbf{X}_t,\varphi) - m(\mathbf{x},\varphi)\big)K\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{X}_t}{h_T}\Big)\,\Big|\,\mathcal{G}_{t-\delta}\Big]\,dt\Big\}. \end{aligned}$$

A simple change of variable combined with the continuity of the conditional density $f^{\mathcal{G}_{t-\delta}}$ and Assumptions (N.2) and (N.6) allow us to infer that

$$\begin{aligned} \mathcal{B}_T(\mathbf{x})| &\le \frac{1}{\widetilde{f}_T(\mathbf{x})}\Big\{\frac{1}{Th_T^d}\int_0^T\int_{\mathbb{R}^d}\big(m(\mathbf{u},\varphi) - m(\mathbf{x},\varphi)\big)K\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{u}}{h_T}\Big)f_t^{\mathcal{G}_{t-\delta}}(\mathbf{u})\,d\mathbf{u}\,dt\Big\} \\ &= \frac{1}{\widetilde{f}_T(\mathbf{x})}\Big\{\frac{1}{T}\int_0^T\int_{\mathbb{R}^d}\big(m(\mathbf{x}+h_T\mathbf{v},\varphi) - m(\mathbf{x},\varphi)\big)K\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{x}}{h_T}+\mathbf{v}\Big)f_t^{\mathcal{G}_{t-\delta}}(\mathbf{x}+h_T\mathbf{v})\,d\mathbf{v}\,dt\Big\} \\ &\le \frac{1}{\delta\widetilde{f}_T(\mathbf{x})}\Big\{\Big(\frac{1}{n}\int_0^T f^{\mathcal{G}_{t-\delta}}(\mathbf{x})\,dt + o(1)\Big)\sup_{\mathbf{x}\in D}\Big|\frac{\partial^r m(\varphi,\mathbf{x})}{\partial v_1^{k_1}\dots\partial v_d^{k_d}}\Big| \\ &\qquad\qquad \times \sum_{k_1+\dots+k_d=r} h_n^{k_1}\dots h_n^{k_d}\int_{[-2L,2L]^d}|v_1^{k_1}\dots v_d^{k_d}|K(\mathbf{0},\mathbf{v})\,d\mathbf{v}\Big\} \\ &\le \frac{1}{\widetilde{f}_T(\mathbf{x})}\big\{c_3h_n^r\big(f(\mathbf{x})+o(1)\big)\big\}. \end{aligned} \tag{6.40}$$

It remains to know that under Assumption (N.6) we obtain

$$\begin{aligned}
\widetilde{f}_T(\mathbf{x}) &= \frac{1}{Th_T^d}\int_0^T \int_{\mathbb{R}^d} K\Big(\frac{\mathbf{x}}{h_T}, \frac{\mathbf{u}}{h_T}\Big) f^{\mathcal{G}_{t-\delta}}(\mathbf{u})\, d\mathbf{u}\, dt \\
&= \frac{1}{T}\int_0^T \int_{[-2L,2L]^d} K\Big(\frac{\mathbf{x}}{h_T}, \frac{\mathbf{x}}{h_T}+\mathbf{v}\Big) f^{\mathcal{G}_{t-\delta}}(\mathbf{x}+h_T\mathbf{v})\, d\mathbf{v}\, dt \\
&= \left(\frac{1}{T}\int_0^T f^{\mathcal{G}_{t-\delta}}(\mathbf{x})\, dt + o(1)\right) \int_{[-2L,2L]^d} K(\mathbf{0},\mathbf{v})\, d\mathbf{v} \\
&= f(\mathbf{x}) + o(1). \qquad (6.41)
\end{aligned}$$

Hence, combining (6.40) and (6.41) we obtain

$$(Th_T^d)^{1/2}\mathcal{B}_T(\mathbf{x}) = O(h_T^r(Th_T^d)^{1/2}). \qquad (6.42)$$

Using a similar argument as for equation (6.8), we may show that

$$\widehat{f}_T(\mathbf{x}) - \widetilde{f}_T(\mathbf{x}) = \big(\frac{\log T}{Th_T}\big)^{1/2} \quad a.s. \qquad (6.43)$$

Combining decomposition (6.38) and equation (6.43) together with (6.42) we get

$$(Th_T^d)^{1/2}\mathcal{R}_T(\mathbf{x}) = O\big(h_T^r(\log T)^{1/2}\big). \qquad (6.44)$$

Making use of Theorem 3.1 and the same steps of the proof implies that, in probability,

$$|\widehat{f}_T(\mathbf{x}) - \mathbb{E}\widehat{f}_T(\mathbf{x})| = o(1). \qquad (6.45)$$

By combining (6.12) and (6.45), we have, in probability, as $n \to \infty$,

$$\widehat{f}_T(\mathbf{x}) \to f(\mathbf{x}). \qquad (6.46)$$

Making use of (6.42), (6.44) and (6.46), we infer that

$$\sqrt{Th_T^d}\big(m_T(\mathbf{x},\varphi) - \overline{m}_T(\mathbf{x},\varphi)\big) = \frac{1}{f(\mathbf{x})}\sqrt{Th_T^d}\,\mathcal{Q}_T(\mathbf{x}) + o_{\mathbb{P}}(1).$$

Consider the decomposition

$$\begin{aligned}
\frac{\sqrt{Th_T^d}}{f(\mathbf{x})}\mathcal{Q}_T(\mathbf{x}) &= \frac{1}{f(\mathbf{x})\sqrt{Th_T^d}}\int_0^T \Big\{\big(\varphi(Y_t) - m(\mathbf{x},\varphi)\big)K\Big(\frac{\mathbf{x}}{h_T}, \frac{\mathbf{X}_t}{h_t}\Big) \\
&\qquad - \mathbb{E}\Big[\big(\varphi(Y_t) - m(\mathbf{x},\varphi)\big)K\Big(\frac{\mathbf{x}}{h_T}, \frac{\mathbf{X}_t}{h_T}\Big)\,\Big|\,\mathcal{G}_{t-\delta}\Big]\Big\}\, dt \\
&= \sum_{i=1}^n \big\{\xi_{T,i}(\mathbf{x},\varphi) - \overline{\xi}_{T,i}(\mathbf{x},\varphi)\big\} = \sum_{i=1}^n \chi_{T,i}(\mathbf{x},\varphi),
\end{aligned}$$

where $T = \delta n$, $i = 1,\ldots,n$,

$$\xi_{T,i}(\mathbf{x},\varphi) = \frac{1}{f(\mathbf{x})\sqrt{Th_T^d}}\int_{(i-1)\delta}^{i\delta} \big(\varphi(Y_t) - m(\mathbf{x},\varphi)\big)K\Big(\frac{\mathbf{x}}{h_T}, \frac{\mathbf{X}_t}{h_t}\Big)\, dt,$$

and

$$\overline{\xi}_{T,i}(\mathbf{x},\varphi) = \frac{1}{f(\mathbf{x})\sqrt{Th_T^d}}\int_{(i-1)\delta}^{i\delta} \mathbb{E}\Big[\big(\varphi(Y_t) - m(\mathbf{x},\varphi)\big)K\Big(\frac{\mathbf{x}}{h_T}, \frac{\mathbf{X}_t}{h_T}\Big)\,\Big|\,\mathcal{G}_{t-\delta}\Big]\, dt,$$

where $\chi_{T,i}(\mathbf{x},\varphi)$ is a triangular array of martingale differences with respect to the $\sigma$-field $\mathcal{G}_i$ (see Didi (2014)). This allows us to apply the central limit theorem for discrete time arrays of martingales (see

Hall and Heyde (1980)) to establish the asymptotic normality of $\sqrt{Th_T^d}\mathcal{Q}_T(\mathbf{x})$. This can be done if we establish the following statements:

(a) Lyapunov's condition:

$$\sum_{i=1}^{n}\mathbb{E}\big[\chi_{T,i}^2(\mathbf{x},\varphi)\mid\mathcal{G}_{i-2}\big]\xrightarrow{\mathbb{P}}\Sigma_\varphi^2(\mathbf{x});$$

(b) Lindeberg's condition:

$$n\mathbb{E}\big[\chi_{T,i}^2(\mathbf{x},\varphi)\mathbf{1}\{|\chi_{T,i}(\mathbf{x},\varphi)|>\epsilon\}\big]=o(1)\quad\text{for any }\ \epsilon>0.$$

**Proof of part (a).** Observe that

$$\left|\sum_{i=1}^{n}\mathbb{E}\big[\xi_{T,i}^2(\mathbf{x},\varphi)\mid\mathcal{G}_{i-2}\big]-\sum_{i=1}^{n}\mathbb{E}\big[\chi_{T,i}^2(\mathbf{x},\varphi)\mid\mathcal{G}_{i-2}\big]\right|=\sum_{i=1}^{n}\big(\mathbb{E}[\xi_{T,i}(\mathbf{x},\varphi)\mid\mathcal{G}_{i-2}]\big)^2.$$

Under Assumptions (N.2) and (N.5), one has

$$\begin{aligned}
&\big|\mathbb{E}[\xi_{T,i}(\mathbf{x},\varphi)\mid\mathcal{G}_{i-2}]\big|\\
&\quad=\frac{1}{f(\mathbf{x})\sqrt{Th_T^d}}\left|\int_{(i-1)\delta}^{i\delta}\mathbb{E}\Big[\big(\varphi(Y_t)-m(\mathbf{x},\varphi)\big)K\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{X}_T}{h_t}\Big)\,\Big|\,\mathcal{G}_{i-2}\Big]\,dt\right|\\
&\quad\le\frac{1}{f(\mathbf{x})\sqrt{Th_T^d}}\left|\int_{(i-1)\delta}^{i\delta}\mathbb{E}\Big[\big(m(\mathbf{X}_t,\varphi)-m(\mathbf{x},\varphi)\big)K\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{X}_t}{h_T}\Big)\,\Big|\,\mathcal{G}_{i-2}\Big]\,dt\right|\\
&\quad=\frac{1}{f(\mathbf{x})\sqrt{Th_T^d}}\int_{(i-1)\delta}^{i\delta}\int_{\mathbb{R}^d}\big(m(\mathbf{u},\varphi)-m(\mathbf{x},\varphi)\big)K\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{u}}{h_T}\Big)f_t^{\mathcal{G}_{i-2}}(\mathbf{u})\,d\mathbf{u}\,dt\\
&\quad\le\frac{\sqrt{h_T}}{f(\mathbf{x})\sqrt{T^d}}\Bigg\{\Big(\int_{(i-1)\delta}^{i\delta}f_t^{\mathcal{G}_{i-2}}(\mathbf{x})\,dt+o(1)\Big)\sup_{\mathbf{x}\in D}\left|\frac{\partial^r m(\varphi,\mathbf{x})}{\partial v_1^{k_1}\dots\partial v_d^{k_d}}\right|\\
&\qquad\qquad\times\sum_{k_1+\dots+k_d=r}h_T^{k_1}\dots h_T^{k_d}\int_{[-2L,2L]^d}|v_1^{k_1}\dots v_d^{k_d}|K(\mathbf{0},\mathbf{v})\,d\mathbf{v}\Bigg\}\\
&\quad\le\frac{c_3h_T^{r+d/2}}{f(\mathbf{x})\sqrt{T}}\Big(\int_{(i-1)\delta}^{i\delta}f_t^{\mathcal{G}_{i-2}}(\mathbf{x})\,dt+o(1)\Big).
\end{aligned}\tag{6.47}$$

Let

$$g_{i-1}^{\mathcal{G}_{i-2}}(\mathbf{x})=\Big(\int_{T_{i-1}}^{T_i}f_t^{\mathcal{G}_{i-2}}(\mathbf{x})\,dt\Big)^2.$$

Making use of the Riemann sum, the quantity $g_{i-1}^{\mathcal{G}_{i-2}}(\mathbf{x})$ may be approximated, whenever $\delta$ is small enough, by $\delta f_{T_{i-1}}^{\mathcal{G}_{i-2}}$. It is then clear from the discussion above that the process $(f_{T_{i-1}}^{\mathcal{G}_{i-2}})_{i\ge1}$ is stationary and ergodic. So the sum $\frac{1}{n}\sum_{i=1}^{n}g_{i-1}^{\mathcal{G}_{i-2}}(\mathbf{x})$ has a finite limit (see Krengel (1985), Theorem 4.4), which is

$$\mathbb{E}\big[g_0^{\mathcal{G}_{-\delta}}(\mathbf{x})\big]=\Big(\int_0^{\delta}f(\mathbf{x})\,dt\Big)^2=\delta^2f^2(\mathbf{x}).\tag{6.48}$$

Therefore by using Condition (6.48) we infer that

$$\sum_{i=1}^{n}\big(\mathbb{E}\big[\xi_{T,i}(\mathbf{x},\varphi)\mid\mathcal{G}_{i-2}\big]\big)^2=\frac{c_3h_n^{2r+d}}{\delta f(\mathbf{x})}\Big(\frac{1}{n}\sum_{i=1}^{n}\Big(\int_{(i-1)\delta}^{i\delta}f_t^{\mathcal{G}_{i-2}}(\mathbf{x})\,dt\Big)^2+o(1)\Big)$$

$$= O(h_T^{2r+d}).$$

The statement (a) follows then from

$$\lim_{n\to\infty}\sum_{i=1}^{n}\mathbb{E}\big[\xi_{T,i}^2(\mathbf{x},\varphi)\mid\mathcal{G}_{i-2}\big]\overset{\mathbb{P}}{=}\Sigma_\varphi^2(\mathbf{x}).$$

Observe that under Condition (N.7) (i) we have

$$\begin{aligned}
&\mathbb{E}\big[\xi_{T,i}^2(\mathbf{x},\varphi)\mid\mathcal{G}_{i-2}\big]\\
&\leq \frac{1}{f^2(\mathbf{x})Th_T^d}\int_{\delta(i-1)}^{i\delta}\mathbb{E}\Big[\mathbb{E}\big[(\varphi(Y_t)-m(\mathbf{x},\varphi))^2\mid\mathcal{S}_{t-\delta,\delta}\big]K^2\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{X}_t}{h_T}\Big)\Bigm|\mathcal{G}_{i-2}\Big]\,dt\\
&= \frac{1}{f^2(\mathbf{x})Th_T^d}\int_{\delta(i-1)}^{i\delta}\mathbb{E}\Big[\sigma_\varphi^2(\mathbf{X}_t)K^2\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{X}_t}{h_T}\Big)\Bigm|\mathcal{G}_{i-2}\Big]\,dt\\
&= \frac{1}{f^2(\mathbf{x})Th_T^d}\int_{\delta(i-1)}^{i\delta}\mathbb{E}\Big[\big(\sigma_\varphi^2(\mathbf{X}_t)-\sigma_\varphi^2(\mathbf{x})\big)K^2\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{X}_t}{h_T}\Big)\Bigm|\mathcal{G}_{i-2}\Big]\,dt\\
&\qquad+\frac{1}{f^2(\mathbf{x})Th_T^d}\int_{\delta(i-1)}^{i\delta}\mathbb{E}\Big[\sigma_\varphi^2(\mathbf{x})K^2\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{X}_i}{h_T}\Big)\Bigm|\mathcal{G}_{i-1}\Big]\,dt\\
&\leq \frac{1}{f^2(\mathbf{x})Th_T^d}\int_{\delta(i-1)}^{i\delta}\mathbb{E}\Big[\sup_{\{\mathbf{u}:\ \|\mathbf{x}-\mathbf{u}\|<h\}}\sup\big|\sigma_\varphi^2(\mathbf{X}_t)-\sigma_\varphi^2(\mathbf{x})\big|K^2\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{X}_t}{h_T}\Big)\Bigm|\mathcal{G}_{i-2}\Big]\,dt\\
&\qquad+\frac{1}{f^2(\mathbf{x})Th_T^d}\int_{\delta(i-1)}^{i\delta}\mathbb{E}\Big[\sigma_\varphi^2(\mathbf{x})K^2\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{X}_t}{h_T}\Big)\Bigm|\mathcal{G}_{i-2}\Big]\,dt\\
&= I_{T,1}(\mathbf{x})+I_{T,2}(\mathbf{x}).
\end{aligned}\tag{6.49}$$

Considering Condition (N.7) (ii) it follows that

$$\begin{aligned}
\Big|\sum_{i=1}^{n}I_{T,1}(\mathbf{x})\Big| &= \frac{1}{f^2(\mathbf{x})Th_T^d}\Big|\sum_{i=1}^{n}\int_{\delta(i-1)}^{i\delta}\mathbb{E}\Big[\sup_{\{\mathbf{u}:\ \|\mathbf{x}-\mathbf{u}\|<h_n\}}\big|\sigma_\varphi^2(\mathbf{X}_t)-\sigma_\varphi^2(\mathbf{x})\big|K^2\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{X}_t}{h_T}\Big)\Bigm|\mathcal{G}_{i-2}\Big]\,dt\Big|\\
&= o(1)\,\frac{1}{f^2(\mathbf{x})}\frac{1}{Th_T^d}\Big|\sum_{i=1}^{n}\int_{\delta(i-1)}^{i\delta}\mathbb{E}\Big[K^2\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{X}_t}{h_T}\Big)\Bigm|\mathcal{G}_{i-2}\Big]\,dt\Big|\\
&= o(1)\,\frac{1}{\delta f^2(\mathbf{x})}\Big|\frac{1}{n}\sum_{i=1}^{n}\int_{\delta(i-1)}^{i\delta}\int_{[-2L,2L]^d}K^2\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{x}}{h_T}+\mathbf{v}\Big)f_t^{\mathcal{G}_{i-2}}(\mathbf{x}+h_T\mathbf{v})\,d\mathbf{v}\,dt\Big|\\
&= o(1)\frac{C_{d+1}}{f^2(\mathbf{x})}\Big|\frac{1}{n}\sum_{i=1}^{n}\int_{\delta(i-1)}^{i\delta}\int_{[-2L,2L]^d}K\Big(\frac{\mathbf{x}}{h_n},\frac{\mathbf{x}}{h_n}+\mathbf{v}\Big)f_t^{\mathcal{G}_{i-2}}(\mathbf{x}+h_n\mathbf{v})\,d\mathbf{v}\,dt\Big|\\
&= o(1)\frac{C_{d+1}}{f^2(\mathbf{x})}\Big|\frac{1}{n}\sum_{i=1}^{n}\Big(\int_{\delta(i-1)}^{i\delta}f_t^{\mathcal{G}_{i-2}}(\mathbf{x})\,dt\Big)+o(1)\Big|\\
&= o(1)\frac{C_{d+1}}{f^2(\mathbf{x})}|f(\mathbf{x})|\\
&= o(1).
\end{aligned}\tag{6.50}$$

Recall that

$$K(\mathbf{u},\mathbf{v})=K(\mathbf{u}+\mathbf{k},\mathbf{v}+\mathbf{k})\qquad\text{for}\quad \mathbf{k}\in\mathbb{Z}^d.$$

Set, for $\mathbf{x}\in D$,

$$\mathbf{x}_m=\Big(\frac{\lfloor 2^m x_1\rfloor}{2^m},\ldots,\frac{\lfloor 2^m x_d\rfloor}{2^m}\Big),$$

with $\lfloor u \rfloor \le u < \lfloor u \rfloor + 1$ denoting the integer part of $u$. Now, we turn our attention to the second term of equation (6.49), it follows then, from Condition (N.6), that

$$\begin{aligned}
\sum_{i=1}^{n} I_{T,2}(\mathbf{x}) &= \frac{1}{f^2(\mathbf{x})Th_T^d}\sum_{i=1}^{n}\int_{T_{i-1}}^{T_i}\mathbb{E}\Big[\sigma_\varphi^2(\mathbf{x})K^2\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{X}_t}{h_T}\Big)\,\Big|\,\mathcal{G}_{i-2}\Big]\,dt \\
&= \frac{\sigma_\varphi^2(\mathbf{x})}{f^2(\mathbf{x})Th_T^d}\sum_{i=1}^{n}\int_{T_{i-1}}^{T_i}\int_{\mathbb{R}^d}K^2\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{u}}{h_T}\Big)f_t^{\mathcal{G}_{i-2}}(\mathbf{u})\,d\mathbf{u}\,dt \\
&= \frac{\sigma_\varphi^2(\mathbf{x})}{f^2(\mathbf{x})Th_T^d}\sum_{i=1}^{n}\int_{T_{i-1}}^{T_i}\int_{\mathbb{R}^d}K^2\Big(\mathbf{0},\frac{\mathbf{u}}{h_T}-\frac{\mathbf{x}_m}{h_T}\Big)f_t^{\mathcal{G}_{i-2}}(\mathbf{u})\,d\mathbf{u}\,dt \\
&+ \frac{\sigma_\varphi^2(\mathbf{x})}{f^2(\mathbf{x})Th_T^d}\sum_{i=1}^{n}\int_{T_{i-1}}^{T_i}\int_{\mathbb{R}^d}\Big\{K^2\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{u}}{h_T}\Big)-K^2\Big(\frac{\mathbf{x}_m}{h_T},\frac{\mathbf{u}}{h_T}\Big)\Big\}f_t^{\mathcal{G}_{i-2}}(\mathbf{u})\,d\mathbf{u}\,dt \\
&= \frac{\sigma_\varphi^2(\mathbf{x})}{\delta n f^2(\mathbf{x})}\sum_{i=1}^{n}\int_{\mathbb{R}^d}K^2(\mathbf{0},\mathbf{v})\Big(\int_{T_{i-1}}^{T_i}f_t^{\mathcal{G}_{i-2}}(\mathbf{x}_m+h_T\mathbf{v})\,dt\Big)\,d\mathbf{v}+o(1) \\
&= \frac{\sigma_\varphi^2(\mathbf{x})}{\delta f^2(\mathbf{x})}\Big(\frac{1}{n}\sum_{i=1}^{n}\int_{T_{i-1}}^{T_i}f_t^{\mathcal{G}_{i-2}}(\mathbf{x})\,dt\Big)\int_{\mathbb{R}^d}K^2(\mathbf{0},\mathbf{v})\,d\mathbf{v}+o(1) \\
&= \frac{\sigma_\varphi^2(\mathbf{x})}{\delta f^2(\mathbf{x})}(\delta f(\mathbf{x})+o(1))\int_{\mathbb{R}^d}\Big\{\sum_{\mathbf{k}\in\mathbb{Z}^d}\phi(\mathbf{k})\phi(\mathbf{v}+\mathbf{k})\Big\}^2\,d\mathbf{v}+o(1) \\
&= \frac{\sigma_\varphi^2(\mathbf{x})}{f(\mathbf{x})}\int_{\mathbb{R}^d}\Big\{\sum_{\mathbf{k}\in\mathbb{Z}^d}\phi(\mathbf{k})\phi(\mathbf{v}+\mathbf{k})\Big\}^2\,d\mathbf{v}+o(1). \qquad (6.51)
\end{aligned}$$

Combining equations (6.50) and (6.51) we get

$$\lim_{n\to\infty}\sum_{i=1}^{n}\mathbb{E}\big[\xi_{T,i}^2(\mathbf{x},\varphi)\mid\mathcal{G}_{i-2}\big]\overset{\mathbb{P}}{=}\Sigma_\varphi^2(\mathbf{x})\le\frac{\sigma_\varphi^2(\mathbf{x})}{f(\mathbf{x})}\int_{\mathbb{R}^d}\Big\{\sum_{\mathbf{k}\in\mathbb{Z}^d}\phi(\mathbf{k})\phi(\mathbf{v}+\mathbf{k})\Big\}^2\,d\mathbf{v}. \qquad (6.52)$$

**Proof of part (b).** The Lindeberg condition results from Corollary 9.5.2 in Chow and Teicher (1997) which implies that

$$n\mathbb{E}\big[\chi_{T,i}^2(\mathbf{x},\varphi)\mathbf{1}\{|\chi_{T,i}(\mathbf{x},\varphi)|>\epsilon\}\big]\le 4n\mathbb{E}\big[\xi_{T,i}^2(\mathbf{x},\varphi)\mathbf{1}\{|\xi_{T,i}(\mathbf{x},\varphi)|>\epsilon/2\}\big].$$

Let $a>1$ and $b>1$ be such that

$$\frac{1}{a}+\frac{1}{b}=1.$$

Making use of Hölder's and Markov's inequalities one can write, for all $\epsilon>0$,

$$\mathbb{E}\big[\xi_{T,i}^2(\mathbf{x},\varphi)\mathbf{1}\{|\xi_{T,i}(\mathbf{x},\varphi)|>\epsilon/2\}\big]\le\frac{\mathbb{E}|\xi_{T,i}(\mathbf{x},\varphi)|^{2a}}{(\epsilon/2)^{2a/b}}.$$

Therefore by using condition (6.2) we obtain

$$\begin{aligned}
&4n\mathbb{E}\big[\xi_{T,i}^2(\mathbf{x},\varphi)\mathbf{1}\{|\xi_{ni}(\mathbf{x},\varphi)|>\epsilon/2\}\big] \\
&\quad\le\frac{4}{\delta T^{a-1}h_T^{ad}(\epsilon/2)^{2a/b}}\mathbb{E}\Big|\int_{T_{i-1}}^{T_i}\big(\varphi(Y_t)-m(\mathbf{x},\varphi)\big)K\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{X}_t}{h_T}\Big)\,dt\Big|^{2a} \\
&\quad\le\frac{4}{\delta T^{a-1}h_T^{(a-1)d}(\epsilon/2)^{2a/b}}\int_{T_{i-1}}^{T_i}\mathbb{E}\Big|\big(\varphi(Y_t)-m(\mathbf{x},\varphi)\big)\frac{C_{d+1}}{h_T^d(1+\|\mathbf{x}-\mathbf{X}_t\|h_T^{-1})^{d+1}}\Big|^{2a}\,dt.
\end{aligned}$$

Using Jensen's and Cauchy–Schwarz's inequalities in combination with Condition (N.0), we infer that

$$4n\mathbb{E}\big[\xi_{T,i}^2(\mathbf{x},\varphi)\mathbf{1}\{|\xi_{ni}(\mathbf{x},\varphi)|>\epsilon/2\}\big] \le \frac{4}{\delta T^{a-1}h_T^{(a-1)d}(\epsilon/2)^{2a/b}}\bigg(\int_{T_{i-1}}^{T_i}\mathbb{E}\big[\big(\varphi(Y_t)-m(\mathbf{x},\varphi)\big)\big]^{2a}\,dt\bigg)$$
$$= O\bigg(\bigg(\frac{1}{Th_T^d}\bigg)^{a-1}\bigg). \tag{6.53}$$

Combining statements (6.52) and (6.53) we achieve the proof of Theorem 4.4. □

*Proof of Theorem 4.5*

Observe that

$$\sqrt{Th_T^d}(\widehat{H}_T(\mathbf{x})-\overline{H}_T(\mathbf{x}))$$
$$=\frac{1}{\sqrt{Th_T^d}}\int_0^T\bigg(\varphi(Y_t)K\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{X}_t}{h_T}\Big)-\mathbb{E}\Big[\varphi(Y_t)K\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{X}_t}{h_T}\Big)\,\Big|\,\mathcal{G}_{t-\delta}\Big]\bigg)\,dt$$
$$=\frac{1}{\sqrt{Th_T^d}}\sum_{i=1}^n\int_{T_{i-1}}^{T_i}\bigg(\varphi(Y_t)K\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{X}_t}{h_T}\Big)-\mathbb{E}\Big[\varphi(Y_t)K\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{X}_t}{h_T}\Big)\,\Big|\,\mathcal{G}_{t-\delta}\Big]\bigg)\,dt$$
$$=\sum_{i=1}^n\big(\xi_{T,i}(\mathbf{x},\varphi)-\widetilde{\xi}_{T,i}(\mathbf{x},\varphi)\big)=\sum_{i=1}^n\chi_{T,i}(\mathbf{x},\varphi),$$

where

$$\chi_{T,i}(\mathbf{x},\varphi)=\frac{1}{\sqrt{Th_T^d}}\int_{T_{i-1}}^{T_i}\bigg(\varphi(Y_t)K\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{X}_t}{h_T}\Big)-\mathbb{E}\Big[\varphi(Y_t)K\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{X}_t}{h_T}\Big)\,\Big|\,\mathcal{G}_{t-\delta}\Big]\bigg)\,dt,$$

is a triangular array of martingale differences with respect to the $\sigma$-field $\mathcal{G}_i$ (see Didi (2014)). Observe that

$$\bigg|\sum_{i=1}^n\mathbb{E}\big[\xi_{T,i}^2(\mathbf{x},\varphi)\mid\mathcal{G}_{i-2}\big]-\sum_{i=1}^n\mathbb{E}\big[\chi_{T,i}^2(\mathbf{x},\varphi)\mid\mathcal{G}_{i-2}\big]\bigg|=\sum_{i=1}^n\big(\mathbb{E}[\xi_{T,i}(\mathbf{x},\varphi)\mid\mathcal{G}_{i-2}]\big)^2.$$

By combining Conditions (N.0)–(N.1) with (6.2), we readily infer that

$$\mathbb{E}\big[\xi_{T,i}(\mathbf{x},\varphi)\mid\mathcal{G}_{i-2}\big]=\frac{1}{\sqrt{Th_T^d}}\int_{T_{i-1}}^{T_i}\mathbb{E}\Big[\varphi(Y_t)K\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{X}_t}{h_T}\Big)\,\Big|\,\mathcal{G}_{i-2}\Big]$$
$$\le\frac{\sqrt{h_T^d}}{\sqrt{T}}\int_{T_{i-1}}^{T_i}\mathbb{E}\Big[\varphi(Y_t)\frac{C_{d+1}}{h_T^d(1+\|\mathbf{x}-\mathbf{X}_t\|h_T^{-1})^{d+1}}\,\Big|\,\mathcal{G}_{i-2}\Big]\,dt$$
$$=\frac{\sqrt{h_T^d}C_{d+1}}{\sqrt{T}}\int_{T_{i-1}}^{T_i}\mathbb{E}[\varphi(Y_t)\mid\mathcal{G}_{i-2}]\,dt,$$

Therefore by Jensen's and Cauchy–Schwarz's inequalities we infer that

$$\sum_{i=1}^n\big(\mathbb{E}[\xi_{T,i}(\mathbf{x},\varphi)\mid\mathcal{G}_{i-2}]\big)^2\le\frac{h_T^dC_{d+1}}{T}\sum_{i=1}^n\bigg(\int_{T_{i-1}}^{T_i}\mathbb{E}\big[\varphi(Y_t)\mid\mathcal{G}_{i-2}\big]\,dt\bigg)^2$$
$$\le\frac{h_T^dC_{d+1}}{T}\sum_{i=1}^n\int_{T_{i-1}}^{T_i}\mathbb{E}\big[\varphi^2(Y_t)\mid\mathcal{G}_{i-2}\big]\,dt$$

$$
\begin{aligned}
&\leq \frac{h_T^d C_{d+1}}{T} \sum_{i=1}^{n} \int_{T_{i-1}}^{T_i} \int_{\mathbb{R}^d} \varphi^2(y) \rho_t^{\mathcal{G}_{i-2}}(y) d\mathbf{v}\, dt \\
&\leq \frac{h_T^d C_{d+1}}{\delta} \left\| \frac{1}{n\rho} \sum_{i=1}^{n} \left( \int_{T_{i-1}}^{T_i} \rho_t^{\mathcal{G}_{i-2}}\, dt \right) \right\| \int_{\mathbb{R}^d} \varphi^2(y) \rho(y)\, d\mathbf{v} \\
&\leq \frac{h_T^d C_{d+1}}{\delta} \left\| \frac{1}{n\rho} \sum_{i=1}^{n} \left( \int_{T_{i-1}}^{T_i} \rho_t^{\mathcal{G}_{i-2}}\, dt \right) \right\| \mathbb{E}[\varphi^2(Y)] \\
&= O(h_T^d).
\end{aligned}
$$

To establish then the asymptotic normality, we have to prove the following statements:

**(a)** Lyapunov's condition:

$$
\sum_{i=1}^{n} \mathbb{E}\big[\xi_{T,i}^2(\mathbf{x},\varphi) \mid \mathcal{G}_{i-2}\big] \xrightarrow{\mathbb{P}} \Sigma^{*2}(\mathbf{x},\varphi);
$$

**(b)** Lindeberg's condition:

$$
n\mathbb{E}\big[\chi_{T,i}^2(\mathbf{x},\varphi)\mathbf{1}\{|\chi_{T,i}(\mathbf{x},\varphi)| > \epsilon\}\big] = o(1) \quad \text{for any } \epsilon > 0.
$$

**Proof of part (a).** Observe that under Assumption (N.2), we obtain

$$
\begin{aligned}
\mathbb{E}\big[\xi_{T,i}^2(\mathbf{x},\varphi) \mid \mathcal{G}_{i-2}\big] &\leq \frac{1}{Th_T^d} \int_{T_{i-1}}^{T_i} \mathbb{E}\left[\varphi^2(Y_t) K^2\Big(\frac{\mathbf{x}}{h_T}, \frac{\mathbf{X}_t}{h_T}\Big) \,\Big|\, \mathcal{G}_{i-2}\right] dt \\
&= \frac{1}{Th_T^d} \int_{T_{i-1}}^{T_i} \mathbb{E}\left[\mathbb{E}\big[\varphi^2(Y_t) \mid \mathcal{S}_{i-1,\delta}\big] K^2\Big(\frac{\mathbf{x}}{h_T}, \frac{\mathbf{X}_t}{h_T}\Big) \,\Big|\, \mathcal{G}_{i-2}\right] dt \\
&= \frac{1}{Th_T^d} \int_{T_{i-1}}^{T_i} \mathbb{E}\left[\mathbb{E}\big[\varphi^2(Y_t) \mid \mathbf{X}_t\big] K^2\Big(\frac{\mathbf{x}}{h_T}, \frac{\mathbf{X}_t}{h_T}\Big) \,\Big|\, \mathcal{G}_{i-2}\right] dt \\
&= \frac{1}{Th_T^d} \int_{T_{i-1}}^{T_i} \mathbb{E}\left[m_2(\varphi,\mathbf{X}_t) K^2\Big(\frac{\mathbf{x}}{h_T}, \frac{\mathbf{X}_t}{h_T}\Big) \,\Big|\, \mathcal{G}_{i-2}\right] dt \\
&= \frac{1}{Th_T^d} \int_{T_{i-1}}^{T_i} \mathbb{E}\left[\big(m_2(\varphi,\mathbf{X}_t) - m_2(\varphi,\mathbf{x})\big) K^2\Big(\frac{\mathbf{x}}{h_T}, \frac{\mathbf{X}_t}{h_T}\Big) \,\Big|\, \mathcal{G}_{i-2}\right] dt \\
&\qquad + \frac{m_2(\varphi,\mathbf{x})}{Th_T^d} \int_{T_{i-1}}^{T_i} \mathbb{E}\left[K^2\Big(\frac{\mathbf{x}}{h_T}, \frac{\mathbf{X}_t}{h_T}\Big) \mid \mathcal{G}_{i-2}\right] dt \\
&= E_{1,i}(\mathbf{x}) + E_{2,i}(\mathbf{x}).
\end{aligned} \tag{6.54}
$$

Using the continuity of the conditional density $f^{\mathcal{G}_{i-1}}(\cdot)$ and Conditions (N.6) and (N.8) and considering a similar argument as for (6.50), we obtain

$$
\begin{aligned}
\sum_{i=1}^{n} E_{1,i}(\mathbf{x}) &= \frac{1}{Th_T^d} \sum_{i=1}^{n} \int_{T_{i-1}}^{T_i} \mathbb{E}\left[\big(m_2(\varphi,\mathbf{X}_t) - m_2(\varphi,\mathbf{x})\big) K^2\Big(\frac{\mathbf{x}}{h_T}, \frac{\mathbf{X}_t}{h_T}\Big) \,\Big|\, \mathcal{G}_{i-2}\right] dt \\
&\leq \frac{1}{Th_T^d} \sum_{i=1}^{n} \int_{T_{i-1}}^{T_i} \mathbb{E}\left[\sup_{\mathbf{u}\in B(\mathbf{x},h_T)} |m_2(\varphi,\mathbf{u}) - m_2(\varphi,\mathbf{x})| K^2\Big(\frac{\mathbf{x}}{h_T}, \frac{\mathbf{X}_t}{h_T}\Big) \,\Big|\, \mathcal{G}_{i-2}\right] dt \\
&\leq \frac{o(1)}{Th_T^d} \sum_{i=1}^{n} \int_{T_{i-1}}^{T_i} \mathbb{E}\left[K^2\Big(\frac{\mathbf{x}}{h_T}, \frac{\mathbf{X}_t}{h_T}\Big) \,\Big|\, \mathcal{G}_{i-2}\right] dt \\
&= \frac{o(1)}{T} \sum_{i=1}^{n} \int_{T_{i-1}}^{T_i} \int_{[-2L,2L]^d} K^2\Big(\frac{\mathbf{x}}{h_T}, \frac{\mathbf{x}}{h_T} + \mathbf{v}\Big) f_t^{\mathcal{G}_{i-2}}(\mathbf{x} + h_T\mathbf{v})\, d\mathbf{v}\, dt
\end{aligned}
$$

$$
\begin{aligned}
&\leq \frac{o(1)C_{d+1}}{\delta n}\sum_{i=1}^{n}\int_{T_{i-1}}^{T_i}\int_{[-2L,2L]^d} K\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{x}}{h_T}+\mathbf{v}\Big) f_t^{\mathcal{G}_{i-2}}(\mathbf{x}+h_T\mathbf{v})\,d\mathbf{v}\,dt\\
&\leq \frac{o(1)C_{d+1}}{\delta}\bigg(\frac{1}{n}\sum_{i=1}^{n}\bigg(\int_{T_{i-1}}^{T_i} f_t^{\mathcal{G}_{i-2}}(\mathbf{x})\,dt\bigg)+o(1)\bigg) = o(1)C_{d+1}f(\mathbf{x})\\
&= o(1).
\end{aligned}
\tag{6.55}
$$

Now, we turn our attention to the second term in (6.54), it follows, from Assumption (N.6), that

$$
\begin{aligned}
\sum_{i=1}^{n} E_{2,i}(\mathbf{x}) &= \frac{1}{Th_T^d}\sum_{i=1}^{n}\int_{T_{i-1}}^{T_i}\mathbb{E}\Big[m_2(\varphi,\mathbf{x})K^2\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{X}_t}{h_T}\Big)\;\Big|\;\mathcal{G}_{i-2}\Big]\,dt\\
&= \frac{m_2(\varphi,\mathbf{x})}{Th_T^d}\sum_{i=1}^{n}\int_{T_{i-1}}^{T_i}\int_{\mathbb{R}^d} K^2\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{u}}{h_T}\Big) f_t^{\mathcal{G}_{i-2}}(\mathbf{u})\,d\mathbf{u}\,dt\\
&= \frac{m_2(\varphi,\mathbf{x})}{Th_T^d}\sum_{i=1}^{n}\int_{T_{i-1}}^{T_i}\int_{\mathbb{R}^d} K^2\Big(\mathbf{0},\frac{\mathbf{u}}{h_T}-\frac{\mathbf{x}_m}{h_T}\Big) f_t^{\mathcal{G}_{i-2}}(\mathbf{u})\,d\mathbf{u}\,dt\\
&\quad+\frac{m_2(\varphi,\mathbf{x})}{Th_T^d}\sum_{i=1}^{n}\int_{T_{i-1}}^{T_i}\int_{\mathbb{R}^d}\Big\{K^2\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{u}}{h_T}\Big)-K^2\Big(\frac{\mathbf{x}_m}{h_T},\frac{\mathbf{u}}{h_T}\Big)\Big\} f_t^{\mathcal{G}_{i-2}}(\mathbf{u})\,d\mathbf{u}\,dt\\
&= \frac{m_2(\varphi,\mathbf{x})}{\delta n}\sum_{i=1}^{n}\int_{\mathbb{R}^d} K^2(\mathbf{0},\mathbf{v})\bigg(\int_{T_{i-1}}^{T_i} f_t^{\mathcal{G}_{i-2}}(\mathbf{x}_m+h_n\mathbf{v})\,dt\bigg)\,d\mathbf{v}+o(1)\\
&= \frac{m_2(\varphi,\mathbf{x})}{\delta}\bigg(\frac{1}{n}\sum_{i=1}^{n}\bigg(\int_{T_{i-1}}^{T_i} f_t^{\mathcal{G}_{i-2}}(\mathbf{x})\,dt\bigg)\bigg)\int_{\mathbb{R}^d} K^2(\mathbf{0},\mathbf{v})\,d\mathbf{v}+o(1)\\
&= m_2(\varphi,\mathbf{x})\big(f(\mathbf{x})+o(1)\big)\int_{\mathbb{R}^d}\Big\{\sum_{\mathbf{k}\in\mathbb{Z}^d}\phi(\mathbf{k})\phi(\mathbf{v}+\mathbf{k})\Big\}^2 d\mathbf{v}+o(1)\\
&= m_2(\varphi,\mathbf{x})f(\mathbf{x})\int_{\mathbb{R}^d}\Big\{\sum_{\mathbf{k}\in\mathbb{Z}^d}\phi(\mathbf{k})\phi(\mathbf{v}+\mathbf{k})\Big\}^2 d\mathbf{v}+o(1).
\end{aligned}
\tag{6.56}
$$

Combining equations (6.55) and (6.56) we get

$$
\lim_{n\to\infty}\sum_{i=1}^{n}\mathbb{E}\big[\xi_{T,i}^2(\mathbf{x},\varphi)\mid\mathcal{G}_{i-2}\big] \overset{\mathbb{P}}{=} \Sigma_\varphi^{*2}(\mathbf{x}) \leq m_2(\varphi,\mathbf{x})f(\mathbf{x})\int_{\mathbb{R}^d}\Big\{\sum_{\mathbf{k}\in\mathbb{Z}^d}\phi(\mathbf{k})\phi(\mathbf{v}+\mathbf{k})\Big\}^2 d\mathbf{v}. \tag{6.57}
$$

**Proof of part (b).** The Lindeberg condition results from Corollary 9.5.2 in Chow and Teicher (1997), which implies that

$$
n\mathbb{E}\big[\chi_{T,i}^2(\mathbf{x},\varphi)\mathbf{1}\{|\chi_{T,i}(\mathbf{x},\varphi)|>\epsilon\}\big] \leq 4n\mathbb{E}\big[\xi_{T,i}^2(\mathbf{x},\varphi)\mathbf{1}\{|\xi_{T,i}(\mathbf{x},\varphi)|>\epsilon/2\}\big].
$$

Let $a>1$ and $b>1$ be such that

$$
\frac{1}{a}+\frac{1}{b}=1.
$$

Making use of Hölder's and Markov's inequalities one can write, for all $\epsilon>0$,

$$
\mathbb{E}\big[\xi_{T,i}^2(\mathbf{x},\varphi)\mathbf{1}\{|\xi_{T,i}(\mathbf{x},\varphi)|>\epsilon/2\}\big] \leq \frac{\mathbb{E}|\xi_{T,i}(\mathbf{x},\varphi)|^{2a}}{(\epsilon/2)^{2a/b}}.
$$

Therefore by Condition (6.2) we have

$$
\begin{aligned}
&4n\mathbb{E}\big[\xi_{T,i}^2(\mathbf{x},\varphi)\mathbf{1}\{|\xi_{T,i}(\mathbf{x},\varphi)|>\epsilon/2\}\big]\\
&\qquad\leq \frac{4}{\delta T^{a-1}h_T^{ad}(\epsilon/2)^{2a/b}}\mathbb{E}\bigg|\int_{T_{i-1}}^{T_i}\varphi(Y_t)K\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{X}_t}{h_T}\Big)\,dt\bigg|^{2a}
\end{aligned}
$$

$$\leq \frac{4}{\delta T^{a-1} h_T^{(a-1)d} (\epsilon/2)^{2a/b}} \int_{T_{i-1}}^{T_i} \mathbb{E}\left|\varphi(Y_t) \frac{C_{d+1}}{h_T^d (1 + \|\mathbf{x} - \mathbf{X}_t\| h_T^{-1})^{d+1}}\right|^{2a} dt.$$

Using Jensen's and Cauchy–Schwarz's inequalities and Condition (N.0), we infer that

$$\begin{aligned}
4n\mathbb{E}\big[\xi_{T,i}^2(\mathbf{x}, \varphi)\mathbf{1}\{|\xi_{T,i}(\mathbf{x}, \varphi)| > \epsilon/2\}\big] &\leq \frac{4}{\delta T^{a-1} h_T^{(a-1)d} (\epsilon/2)^{2a/b}} \left( \int_{T_{i-1}}^{T_i} \mathbb{E}[\varphi^{2a}(Y_t)]\, dt \right) \\
&= O\left( \left( \frac{1}{T h_T^d} \right)^{a-1} \right). \qquad (6.58)
\end{aligned}$$

Combining statements (6.57) and (6.58) we achieve the proof of the theorem. □

*Proof of Corollary 4.6*

Consider the following decomposition

$$\begin{aligned}
\sqrt{T h_T^d}\big(\widehat{H}_T(\mathbf{x}) - H(\mathbf{x})\big) &= \sqrt{T h_T^d}\big(\widehat{H}_T(\mathbf{x}) - \overline{H}_T(\mathbf{x})\big) + \sqrt{T h_T^d}\big(\overline{H}_T(\mathbf{x}) - \mathbb{E}(\widehat{H}_T(\mathbf{x}))\big) \\
&\quad + \sqrt{T h_T^d}\big(\mathbb{E}(\widehat{H}_T(\mathbf{x})) - H(\mathbf{x})\big) \\
&= U_T(\mathbf{x}, \varphi) + W_T(\mathbf{x}, \varphi) + V_T(\mathbf{x}, \varphi).
\end{aligned}$$

Concerning the convergence in distribution of the term $U_T(\mathbf{x}, \varphi)$, the result follows from Theorem 4.5. On the other hand, observe that

$$\begin{aligned}
\frac{1}{\sqrt{T h_T^d}} W_T(\mathbf{x}, \varphi) &= \overline{H}_T(\mathbf{x}) - \mathbb{E}\big(\widehat{H}_T(\mathbf{x})\big) \\
&= \frac{1}{T h_T^d} \int_0^T \left( \mathbb{E}\Big[\varphi(Y_t) K\Big(\frac{\mathbf{x}}{h_T}, \frac{\mathbf{X}_t}{h_T}\Big) \,\Big|\, \mathcal{G}_{t-\delta}\Big] - \mathbb{E}\Big[\varphi(Y_t) K\Big(\frac{\mathbf{x}}{h_T}, \frac{\mathbf{X}_t}{h_T}\Big)\Big] \right) dt \\
&= \frac{1}{T h_T^d} \int_0^T \int_{\mathbb{R}^{d+1}} \varphi(y) K\Big(\frac{\mathbf{x}}{h_T}, \frac{\mathbf{u}}{h_T}\Big) \big(g_t^{\mathcal{G}_{t-\delta}}(\mathbf{u}, y) - g(\mathbf{u}, y)\big)\, d\mathbf{u}\, dy\, dt \\
&\leq \frac{1}{T} \int_0^T \int_{\mathbb{R}^{d+1}} \varphi(y) \frac{C_{d+1}}{h_T^d (1 + \|\mathbf{x} - \mathbf{u}\| h_T^{-1})^{d+1}} \big(g_t^{\mathcal{G}_{t-\delta}}(\mathbf{u}, y) - g(\mathbf{u}, y)\big)\, d\mathbf{u}\, dy\, dt \\
&= h_T C_{d+1} \int_{\mathbb{R}^{d+1}} \varphi(y) \left( T^{-1} \int_0^T g_t^{\mathcal{G}_{t-\delta}}(\mathbf{u}, y)\, dt - g(\mathbf{u}, y) \right) \mathbf{u}\, dy.
\end{aligned}$$

The Cauchy–Schwarz inequality combined with Assumptions (N.0) and (N.5) implies that

$$\frac{1}{\sqrt{T h_T^d}} W_T(\mathbf{x}, \varphi) \leq h_T C_{d+1} \|\varphi\|_{L^2} \left\| T^{-1} \int_0^T g_t^{\mathcal{G}_{t-\delta}}\, dt - g \right\|_{L^2} = O(h_T).$$

Condition (4.8) completes the proof. By change of variables, it follows that

$$\begin{aligned}
\frac{1}{\sqrt{T h_T^d}} V_T(\mathbf{x}, \varphi) &= \big|\mathbb{E}\big(\widehat{H}_T(\mathbf{x})\big) - H(\mathbf{x})\big| \\
&= \frac{1}{T h_T^d} \int_0^T \mathbb{E}\Big[\varphi(Y_t) K\Big(\frac{\mathbf{x}}{h_T}, \frac{\mathbf{X}_t}{h_T}\Big)\Big]\, dt - H(\mathbf{x}) \\
&= \frac{1}{T h_T^d} \int_0^T \mathbb{E}\Big[\mathbb{E}\big[\varphi(Y_t) \mid \mathbf{X}_t\big] K\Big(\frac{\mathbf{x}}{h_T}, \frac{\mathbf{X}_t}{h_T}\Big)\Big]\, dt - H(\mathbf{x}) \\
&= \frac{1}{T h_T^d} \int_0^T \mathbb{E}\Big[m(\varphi, \mathbf{X}_t) K\Big(\frac{\mathbf{x}}{h_T}, \frac{\mathbf{X}_t}{h_T}\Big)\Big]\, dt - m(\varphi, \mathbf{x}) f(\mathbf{x})
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{h_T^d}\int_{\mathbb{R}^d} m(\varphi,\mathbf{u})K\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{u}}{h_T}\Big)f(\mathbf{u})\,d\mathbf{u} - m(\varphi,\mathbf{x})f(\mathbf{x})\\
&= \int_{[-2L,2L]^d} m(\varphi,\mathbf{x}+h_T\mathbf{v})f(\mathbf{x}+h_T\mathbf{v})K\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{x}}{h_T}+\mathbf{v}\Big)\,d\mathbf{v} - m(\varphi,\mathbf{x})f(\mathbf{x})\\
&= \int_{[-2L,2L]^d} (m(\varphi,\mathbf{x}+h_T\mathbf{v})f(\mathbf{x}+h_T\mathbf{v}) - m(\varphi,\mathbf{x})f(\mathbf{x}))K\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{x}}{h_T}+\mathbf{v}\Big)\,d\mathbf{v}.
\end{aligned}
$$

Making use of Conditions (C.3), (N.4) and using the same previous arguments, we infer that

$$
\begin{aligned}
|V_T(\mathbf{x},\varphi)| &\le \frac{\sqrt{Th_T^d}}{r!}\sup_{\mathbf{x}\in D}\left|\frac{\partial^r m(\varphi,\mathbf{x})f(\mathbf{x})}{\partial v_1^{k_1}\dots\partial v_d^{k_d}}\right|\\
&\quad\times\sum_{k_1+\dots+k_d=r} h_T^{k_1}\dots h_T^{k_d}\int_{[-2L,2L]^d}|v_1^{k_1}\dots v_d^{k_d}|K\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{x}}{h_T}+\mathbf{v}\Big)\,d\mathbf{v}\\
&= O\Big(h_T^r\Big(\sqrt{Th_T^d}\Big)\Big).
\end{aligned}
$$

Therefore

$$
V_T(\mathbf{x},\varphi) = \sqrt{Th_T^d}\big(\mathbb{E}(\widehat{H}_T(\mathbf{x})) - H(\mathbf{x})\big) = O\Big(h_T^r\Big(\sqrt{Th_T^d}\Big)\Big).
$$

We achieve then the proof of the Corollary. □

*Proof of Theorem 4.7*

Recall that

$$
\begin{aligned}
\sqrt{Th_T^d}\big(\widehat{f}_T(\mathbf{x}) - \overline{f}(\mathbf{x})\big) &= \frac{1}{\sqrt{Th_T^d}}\int_0^T\left(K\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{X}_t}{h_T}\Big) - \mathbb{E}\Big[K\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{X}_t}{h_T}\Big)\,\Big|\,\mathcal{F}_{t-\delta}\Big]\right)dt\\
&= \sum_{i=1}^n\big(\eta_{T,i}(\mathbf{x}) - \widetilde{\eta}_{T,i}(\mathbf{x})\big) = \sum_{i=1}^n\theta_{T,i}(\mathbf{x}),
\end{aligned}
$$

where

$$
\begin{aligned}
\eta_{T,i}(\mathbf{x}) &= \frac{1}{\sqrt{Th_T^d}}\int_{T_{i-1}}^{T_i}K\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{X}_t}{h_T}\Big)\,dt,\\
\widetilde{\eta}_{T,i}(\mathbf{x}) &= \frac{1}{\sqrt{Th_T^d}}\int_{T_{i-1}}^{T_i}\mathbb{E}\Big[K\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{X}_t}{h_T}\Big)\,\Big|\,\mathcal{F}_{t-\delta}\Big]\,dt,\\
\theta_{T,i}(\mathbf{x}) &= \big(\eta_{T,i}(\mathbf{x}) - \widetilde{\eta}_{T,i}(\mathbf{x})\big).
\end{aligned}
$$

The sequence $\{\theta_{T,i}(\mathbf{x})\}$ is a triangular array of martingale differences with respect to the $\sigma$-field $\mathcal{F}_i$ (see Didi (2014)). Observe that

$$
\left|\sum_{i=1}^n\mathbb{E}\big[\eta_{T,i}^2(\mathbf{x})\mid\mathcal{F}_{i-2}\big] - \sum_{i=1}^n\mathbb{E}\big[\theta_{T,i}^2(\mathbf{x})\mid\mathcal{F}_{i-2}\big]\right| = \sum_{i=1}^n\big(\mathbb{E}\big[\eta_{T,i}(\mathbf{x})\mid\mathcal{F}_{i-2}\big]\big)^2.
$$

By a simple change of variable, we obtain

$$
\mathbb{E}\big[\eta_{T,i}(\mathbf{x})\mid\mathcal{F}_{i-2}\big] = \frac{1}{\sqrt{Th_T^d}}\int_{T_{i-1}}^{T_i}\mathbb{E}\Big[K\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{X}_t}{h_T}\Big)\,\Big|\,\mathcal{F}_{i-2}\Big]\,dt
$$

$$= \frac{\sqrt{h_T^d}}{\sqrt{T}} \int_{T_{i-1}}^{T_i} \int_{\mathbb{R}^d} K\Big(\frac{\mathbf{x}}{h_T}, \frac{\mathbf{u}}{h_T}\Big) f_t^{\mathcal{F}_{i-2}}(\mathbf{u})\, d\mathbf{u}\, dt$$
$$= \frac{\sqrt{h_T^d}}{\sqrt{T}} \int_{[-2L,2L]^d} K\Big(\frac{\mathbf{x}}{h_T}, \frac{\mathbf{x}}{h_T} + \mathbf{v}\Big) f_t^{\mathcal{F}_{i-2}}(\mathbf{x} + h_T\mathbf{v})\, d\mathbf{v}\, dt$$
$$= \frac{\sqrt{h_T^d}}{\sqrt{T}} \Big( \int_{T_{i-1}}^{T_i} f_t^{\mathcal{F}_{i-2}}(\mathbf{x})\, dt + o(1) \Big) \quad a.s.$$

Considering Assumption (C.2) and recalling that the function $\mathbf{x} \mapsto \mathbf{x}^2$ is measurable we have

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \Big( \int_{T_{i-1}}^{T_i} f_t^{\mathcal{F}_{i-2}}(\mathbf{x})\, dt \Big)^2 = \delta^2 f^2(\mathbf{x}) \qquad \text{in the } a.s. \text{ and } L^2 \text{ sense.}$$

Hence we obtain

$$\sum_{i=1}^{n} \big(\mathbb{E}[\eta_{T,i}(\mathbf{x},\varphi) \mid \mathcal{F}_{i-2}]\big)^2 \le \frac{h_T^d}{T} \sum_{i=1}^{n} \Big( \int_{T_{i-1}}^{T_i} f_t^{\mathcal{F}_{i-2}}(\mathbf{x})\, dt + o(1) \Big)^2$$
$$\le \frac{h_T^d}{\delta n} \sum_{i=1}^{n} \Big( \int_{T_{i-1}}^{T_i} f_t^{\mathcal{F}_{i-2}}(\mathbf{x}) \Big)^2 + o(h_T^d) \Big( \frac{1}{n} \sum_{i=1}^{n} \Big( \int_{T_{i-1}}^{T_i} f_t^{\mathcal{F}_{i-2}}(\mathbf{x})\, dt \Big) \Big) + o(h_T^{2d})$$
$$= O(h_T^d).$$

To establish the asymptotic normality of $\sqrt{T h_T^d}\big(\widehat{f}_T(\mathbf{x}) - \overline{f}(\mathbf{x})\big)$ we have to show the following statements:

**(a)** Lyapunov's condition:

$$\sum_{i=1}^{n} \mathbb{E}\big[\eta_{T,i}^2(\mathbf{x}) \mid \mathcal{F}_{i-2}\big] \xrightarrow{\mathbb{P}} \Sigma^{**2}(\mathbf{x});$$

**(b)** Lindeberg's condition:

$$n\mathbb{E}\big[\theta_{T,i}^2(\mathbf{x}) \mathbf{1}\{|\theta_{T,i}(\mathbf{x})| > \epsilon\}\big] = o(1) \quad \text{for any } \epsilon > 0.$$

**Proof of part (a).** Using a change of variable and the continuity of the conditional density $f_t^{\mathcal{F}_{i-2}}(\cdot)$, $t \in [T_{i-1}, T_i]$, we obtain

$$\mathbb{E}\big[\eta_{T,i}^2(\mathbf{x}) \mid \mathcal{F}_{i-2}\big] \le \frac{1}{T h_T^d} \int_{T_{i-1}}^{T_i} \mathbb{E}\Big[K^2\Big(\frac{\mathbf{x}}{h_T}, \frac{\mathbf{X}_t}{h_T}\Big) \Big| \mathcal{F}_{i-2}\Big]\, dt$$
$$= \frac{1}{\delta n} \int_{T_{i-1}}^{T_i} \int_{[-2L,2L]^d} K^2\Big(\frac{\mathbf{x}}{h_T}, \frac{\mathbf{x}}{h_T} + \mathbf{v}\Big) f_t^{\mathcal{F}_{i-2}}(\mathbf{x} + h_T\mathbf{v})\, d\mathbf{v}\, dt$$
$$= \frac{1}{\delta n} \Big( \int_{T_{i-1}}^{T_i} f_t^{\mathcal{F}_{i-2}}(\mathbf{x})\, dt + o(1) \Big) \int_{[-2L,2L]^d} K^2\Big(\frac{\mathbf{x}}{h_T}, \frac{\mathbf{x}}{h_T} + \mathbf{v}\Big)\, d\mathbf{v}.$$

Recall that

$$K(\mathbf{u},\mathbf{v}) = K(\mathbf{u}+\mathbf{k}, \mathbf{v}+\mathbf{k}) \quad \text{for } \mathbf{k} \in \mathbb{Z}^d.$$

Moreover, the ergodicity of the process $\{f_t^{\mathcal{F}_{t-\delta}}\}$ implies the ergodicity of $\{f_t^{\mathcal{F}_{T_{i-2}}}\}$ for $t \in [T_{i-1}, T_i]$. We obtain then

$$\frac{1}{n} \sum_{i=1}^{n} \Big( \frac{1}{\delta} \int_{T_{i-1}}^{T_i} f_t^{\mathcal{F}_{i-2}}(\mathbf{x})\, dt \Big) \underset{T\to\infty}{=} f(\mathbf{x}) \quad \text{in} \quad a.s. \quad \text{and} \quad L^2 \text{ sense.}$$

Assumption (C.2) allows us to write

$$\begin{aligned}\sum_{i=1}^{n}\mathbb{E}\big[\eta_{T,i}^2(\mathbf{x})\mid\mathcal{F}_{i-1}\big] &\leq \left(\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\delta}\int_{T_{i-1}}^{T_i} f_t^{\mathcal{F}_{i-2}}(\mathbf{x})\,dt\right)+o(1)\right)\int_{[-2L,2L]^d}K^2(0,\mathbf{v})\,d\mathbf{v}\\ &= f(\mathbf{x})\int_{[-2L,2L]^d}K^2(0,\mathbf{v})\,d\mathbf{v}+o(1)\\ &= f(\mathbf{x})\int_{\mathbb{R}^d}\left\{\sum_{\mathbf{k}\in\mathbb{Z}^d}\phi(\mathbf{k})\phi(\mathbf{v}+\mathbf{k})\right\}^2 d\mathbf{v}+o(1).\end{aligned}$$

Therefore

$$\lim_{n\to\infty}\sum_{i=1}^{n}\mathbb{E}\big[\eta_{T,i}^2(\mathbf{x})\mid\mathcal{F}_{i-2}\big]\overset{\mathbb{P}}{=}\Sigma^{**2}(\mathbf{x})\leq f(\mathbf{x})\int_{\mathbb{R}^d}\left\{\sum_{\mathbf{k}\in\mathbb{Z}^d}\phi(\mathbf{k})\phi(\mathbf{v}+\mathbf{k})\right\}^2 d\mathbf{v}. \tag{6.59}$$

**Proof of part (b).** We observe that

$$n\mathbb{E}\big[\theta_{T,i}^2(\mathbf{x})\mathbf{1}\{|\theta_{T,i}(\mathbf{x})|>\epsilon\}\big]\leq 4n\mathbb{E}\big[\eta_{T,i}^2(\mathbf{x})\mathbf{1}\{|\eta_{T,i}(\mathbf{x})|>\epsilon/2\}\big].$$

Let $a>1$ and $b>1$ such that

$$\frac{1}{a}+\frac{1}{b}=1.$$

Making use of Hölder's and Markov's inequalities one can write, for all $\epsilon>0$,

$$\mathbb{E}\big[\eta_{T,i}^2(\mathbf{x})\mathbf{1}\{|\eta_{T,i}(\mathbf{x})|>\epsilon/2\}\big]\leq\frac{\mathbb{E}|\eta_{T,i}(\mathbf{x})|^{2a}}{(\epsilon/2)^{2a/b}}.$$

Therefore by Jensen's inequality

$$\begin{aligned}&4n\mathbb{E}\big[\eta_{T,i}^2(\mathbf{x})\mathbf{1}\{|\eta_{T,i}(\mathbf{x})|>\epsilon/2\}\big]\\ &\leq\frac{4}{\delta T^{a-1}h_T^{ad}(\epsilon/2)^{2a/b}}\int_{T_{i-1}}^{T_i}\mathbb{E}\left|K\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{X}_t}{h_T}\Big)\right|^{2a}dt\\ &=\frac{4}{\delta(Th_T^d)^{a-1}(\epsilon/2)^{2a/b}}\int_{T_{i-1}}^{T_i}\int_{[-2L,2L]^d}K^{2a}\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{x}}{h_T}+\mathbf{v}\Big)f(\mathbf{x}+h_T\mathbf{v})\,d\mathbf{v}\,dt\\ &=\frac{4}{(Th_T^d)^{a-1}(\epsilon/2)^{2a/b}}\big(f(\mathbf{x})+o(1)\big)\int_{[-2L,2L]^d}K^{2a}\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{x}}{h_T}+\mathbf{v}\Big)\,d\mathbf{v}\\ &=O\Big(\Big(\frac{1}{nh_n^d}\Big)^{a-1}\Big).\end{aligned} \tag{6.60}$$

Combining statements (6.59) and (6.60) we complete the proof of Theorem 4.7. □

*Proof of Corollary 4.8*

Consider the following decomposition

$$\begin{aligned}\sqrt{Th_T^d}\big(\widehat{f}_T(\mathbf{x})-f(\mathbf{x})\big)&=\sqrt{Th_T^d}\big(\widehat{f}_T(\mathbf{x})-\overline{f}_T(\mathbf{x})\big)+\sqrt{Th_T^d}\big(\overline{f}_T(\mathbf{x})-\mathbb{E}(\widehat{f}_T(\mathbf{x}))\big)\\ &\quad+\sqrt{Th_T^d}\big(\mathbb{E}(\widehat{f}_T(\mathbf{x}))-f(\mathbf{x})\big)\\ &=F_{T,1}(\mathbf{x})+F_{T,2}(\mathbf{x})+F_{T,3}(\mathbf{x}).\end{aligned}$$

Theorem 4.7 shows the convergence in distribution of $F_{T,1}(\mathbf{x})$ to a normal distribution. Secondly, by combining Condition (C.1) with (6.2), we obtain

$$\frac{1}{\sqrt{Th_T^d}}F_{T,2}(\mathbf{x},\varphi)=\overline{f}_T(\mathbf{x})-\mathbb{E}\big(\widehat{f}_T(\mathbf{x})\big)$$

$$
\begin{aligned}
&= \frac{1}{Th_T^d}\int_0^T \left(\mathbb{E}\Big[K\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{X}_t}{h_T}\Big)\,\Big|\,\mathcal{F}_{t-\delta}\Big] - \mathbb{E}\Big[K\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{X}_t}{h_T}\Big)\Big]\right)dt\\
&= \frac{1}{Th_T^d}\int_0^T\int_{\mathbb{R}^d} K\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{u}}{h_T}\Big)\big(f_t^{\mathcal{F}_{t-\delta}}(\mathbf{u}) - f(\mathbf{u})\big)\,d\mathbf{u}\,dt\\
&\le \frac{1}{T}\int_0^T\int_{\mathbb{R}^d} \frac{C_{d+1}}{h_T^d(1+\|\mathbf{x}-\mathbf{u}\|h_T^{-1})^{d+1}}\big(f_t^{\mathcal{F}_{t-\delta}}(\mathbf{u}) - f(\mathbf{u})\big)\,d\mathbf{u}\,dt\\
&= h_T C_{d+1}\int_{\mathbb{R}^d}\left(T^{-1}\int_0^T f_t^{\mathcal{F}_{t-\delta}}(\mathbf{u})\,dt - f(\mathbf{u})\right)d\mathbf{u}.
\end{aligned}
$$

Using the fact that

$$
\int_{\mathbb{R}^d} f^{\mathcal{F}_{t-\delta}}(\mathbf{u})\,d\mathbf{u} = 1 \quad \text{and} \quad \int_{\mathbb{R}^d} f(\mathbf{u})\,d\mathbf{u} = 1,
$$

we infer that

$$
\frac{1}{\sqrt{Th_T^d}}F_{T,2}(\mathbf{x}) \le 2h_T C_{d+1} = O(h_T).
$$

Condition (4.9) implies readily that

$$
\lim_{T\to\infty} F_{T,2}(\mathbf{x}) = 0.
$$

Under assumption (C.3), we get

$$
\begin{aligned}
F_{T,3}(\mathbf{x}) &= \sqrt{Th_T^d}|\mathbb{E}\widehat{f}_T(\mathbf{x}) - f(\mathbf{x})|\\
&= \left|\sqrt{Th_T^d}\frac{1}{h_T^d}\int_{\mathbb{R}^d} K\left(\frac{\mathbf{x}}{h_T},\frac{\mathbf{v}}{h_T}\right)f(\mathbf{v})\,d\mathbf{v} - f(\mathbf{x})\right|\\
&= \left|\sqrt{Th_T^d}\int_{[-2L,2L]^d} K\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{x}}{h_T}+\mathbf{v}\Big)\big(f(\mathbf{x}+h_T\mathbf{v}) - f(\mathbf{x})\big)\,d\mathbf{v}\right|\\
&= \left|\sqrt{Th_T^d}\int_{[-2L,2L]^d} K\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{x}}{h_T}+\mathbf{v}\Big)\left(\frac{1}{r!}\sum_{k_1+\cdots+k_d=r} h_T^{k_1}v_1^{k_1}\dots h_T^{k_d}v_d^{k_d}\frac{\partial^r f(\mathbf{v}h_T\theta+\mathbf{x})}{\partial v_1^{k_1}\dots\partial v_d^{k_d}}\right)d\mathbf{v}\right|\\
&\le \frac{\sqrt{Th_T^d}}{r!}\sup_{\mathbf{x}\in D}\left|\frac{\partial^r f(\mathbf{x})}{\partial v_1^{k_1}\dots\partial v_d^{k_d}}\right|\sum_{k_1+\cdots+k_d=r} h_T^{k_1}\dots h_T^{k_d}\int_{[-2L,2L]^d}|v_1^{k_1}\dots v_d^{k_d}|K\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{x}}{h_T}+\mathbf{v}\Big)\,d\mathbf{v}\\
&= \frac{\sqrt{Th_T^d}}{r!}\sup_{\mathbf{x}\in D}\left|\frac{\partial^r f(\mathbf{x})}{\partial v_1^{k_1}\dots\partial v_d^{k_d}}\right|h_T^r\sum_{k_1+\cdots+k_d=r}\int_{[-2L,2L]^d}|v_1^{k_1}\dots v_d^{k_d}|\left|K\Big(\frac{\mathbf{x}}{h_T},\frac{\mathbf{x}}{h_T}+\mathbf{v}\Big)\right|d\mathbf{v}\\
&= O\Big(h_T^r\Big(\sqrt{Th_T^d}\Big)\Big).
\end{aligned}
$$

Therefore

$$
\lim_{T\to\infty} F_{T,3}(\mathbf{x}) = 0.
$$

The proof of Corollary 4.8 is completed. □

## ACKNOWLEDGEMENTS

The authors are indebted to the Editor and the referee for their constructive criticism, very valuable comments and suggestions which led to a considerable improvement of the manuscript. The authors also thank their colleague Professor Djalil KATEB for his help and discussions.

## REFERENCES


1. D. W. K. Andrews, "Nonstrong Mixing Autoregressive Processes", J. Appl. Probab. **21** (4), 930–934 (1984).
2. G. Banon, "Nonparametric Identification for Diffusion Processes", SIAM J. Control Optim. **16** (3), 380–395 (1978).
3. A. Beck, "On the Strong Law of Large Numbers", in *Ergodic Theory (Proc. Internat. Sympos., Tulane Univ., New Orleans, La., 1961)*, (Academic Press, New York, 1963), pp 21–53.
4. J.Bergh and J. Löfström, *Interpolation Spaces. An Introduction*, in *Grundlehren der Mathematischen Wissenschaften* (Springer-Verlag, Berlin, 1976), No. 223.
5. D. Bosq and J.-P. Lecoutre, *Théorie de l'estimation fonctionnelle*, in *Economie et Statistiques Avancées* (Economica, Paris, 1987).
6. S. Bouzebda and S. Didi, "Multivariate Wavelet Density and Regression Estimators for Stationary and Ergodic Discrete Time Processes: Asymptotic Results", Comm. Statist. Theory Methods, (2015) (in press). DOI:10.1080/03610926.2015.1019144
7. R. C. Bradley, *Introduction to Strong Mixing Conditions* (Kendrick Press, Heber City, UT, 2007), Vol. 1.
8. J. E. Chacón, J. Montanero, and A. G. Nogales, "A Note on Kernel Density Estimation at a Parametric Rate", J. Nonparam. Statist. **19** (1), 13–21 (2007).
9. Y. S. Chow and H. Teicher, *Probability Theory* in *Springer Texts in Statistics*, (Springer, New York, 1997).
10. I. Daubechies, *Ten Lectures on Wavelets*, in *CBMS-NSF Regional Conf. Ser. in Appl. Math.* (SIAM, Philadelphia, PA, 1992), Vol. 61.
11. V. H. de la Peña and E. Giné, *Decoupling. From Dependence to Independence, Randomly Stopped Processes. U-Statistics and Processes. Martingales and Beyond*, in *Probability and Its Applications* (Springer-Verlag, New York, 1999).
12. M. Delecroix, *Sur l'estimation et la prévision non-paramétrique des processus ergodiques*, Doctorat d'État. Univ. des Sci. de Lille, Flandre-Artois (1987).
13. L. Devroye and L. Györfi, *Nonparametric Density Estimation. The $L_1$ View*, in *Wiley Series in Probability and Mathematical Statistics: Tracts on Probability and Statistics* (Wiley, New York, 1985).
14. L. Devroye and G. Lugosi, *Combinatorial Methods in Density Estimation*, in *Springer Series in Statistics* (Springer-Verlag, New York, 2001).
15. S. Didi, *Quelques propriétés asymptotiques en estimation non paramétrique de fonctionnelles de processus stationnaires à temps continu*, Thèse Doctorat. Univ. Pierre et Marie Curie (2014).
16. E. Giné and W. R. Madych, "On Wavelet Projection Kernels and the Integrated Squared Error in Density Estimation", Statist. Probab. Lett. **91**, 32–40 (2014).
17. E. Giné and R. Nickl, "Uniform Limit Theorems for Wavelet Density Estimators", Ann. Probab. **37** (4), 1605–1646 (2009).
18. L. Györfi, W. Härdle, P. Sarda, and P. Vieu, *Nonparametric Curve Estimation from Time Series*, in *Lecture Notes in Statistics* (Springer-Verlag, Berlin, 1989), Vol. 60.
19. P. Hall and C. C. Heyde, C. C. *Martingale Limit Theory and Its Application*, in *Probability and Mathematical Statistics* (Academic Press, New York, 1980).
20. W. Härdle, G. Kerkyacharian, D. Picard, and A. Tsybakov, *Wavelets, Approximation, and Statistical Applications*, in *Lecture Notes in Statistics* (Springer-Verlag, New York, 1998), Vol. 129.
21. M. C. Jones, "On Higher Order Kernels", J. Nonparam. Statist. **5** (2), 215–221 (1995).
22. M. C. Jones and D. F. Signorini, "A Comparison of Higher-Order Bias Kernel Density Estimators", J. Amer. Statist. Assoc. **92** (439), 1063–1073 (1997).
23. M. C. Jones, S. J. Davies, and B. U. Park, "Versions of Kernel-Type Regression Estimators", J. Amer. Statist. Assoc. **89** (427), 825–832 (1994).
24. M. C. Jones, O. Linton, and J. P. Nielsen, "A Simple Bias Reduction Method for Density Estimation", Biometrika **82** (2), 327–338 (1995).
25. U. Krengel, *Ergodic Theorems*, in *de Gruyter Studies in Mathematics* (de Gruyter, Berlin, 1985), Vol. 6. With a supplement by Antoine Brunel.
26. F. Leblanc, "Discretized Wavelet Density Estimators for Continuous Time Stochastic Processes", in *Lecture Notes in Statist.*, Vol. 103: *Wavelets and Statistics (Villard de Lans, 1994)* (Springer, New York, 1995), pp. 209–224.
27. A. Leucht and M. H. Neumann, "Degenerate $U$- and $V$-Statistics under Ergodicity: Asymptotics, Bootstrap and Applications in Statistics", Ann. Inst. Statist. Math. **65**, 349–386 (2013).
28. L. Li, "Nonparametric Adaptive Density Estimation on Random Fields Using Wavelet Method", Statist. Probab. Lett. **96**, 346–355 (2014).
29. S. Mallat, *A Wavelet Tour of Signal Processing. The Sparse Way*, 3ed ed. (Elsevier/Academic Press, Amsterdam, 2009). With contributions from Gabriel Peyré.
30. E. Masry, "Multivariate Probability Density Estimation by Wavelet Methods: Strong Consistency and Rates for Stationary Time Series", Stochastic Process. Appl. **67** (2), 177–193 (1997).

31. E. Masry, "Wavelet-Based Estimation of Multivariate Regression Functions in Besov Spaces", J. Nonparam. Statist. **12** (2), 283–308 (2000).
32. Y. Meyer, *Wavelets and Operators*, in *Cambridge Studies in Advanced Mathematics* (Cambridge Univ. Press, Cambridge, 1992), Vol. 37. Transl. from 1990 French original by D. H. Salinger.
33. M. H. Neumann, "Absolute Regularity and Ergodicity of Poisson Count Processes", Bernoulli **17** (4), 1268–1284 (2011).
34. B. Vidakovic, *Statistical Modeling by Wavelets*, in *Wiley Series in Probability and Statistics: Applied Probability and Statistics* (Wiley, New York, 1999).
35. H. Triebel, *Theory of Function Spaces*, in *Monographs in Mathematics* (Birkhäuser, Basel, 1983), Vol. 78.
36. B. L. S. Prakasa Rao, *Nonparametric Functional Estimation*, in *Probability and Mathematical Statistics* (Academic Press, New York, 1983).
37. G. Peškir, "The Uniform Mean-Square Ergodic Theorem for Wide Sense Stationary Processes", Stochastic Anal. Appl. **16** (4), 697–720 (1998).
38. D. W. Scott, *Multivariate Density Estimation. Theory, Practice, and Visualization*, in *Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics* (Wiley, New York, 1992).
39. M. P. Wand and M. C. Jones, *Kernel Smoothing*, in *Monographs on Statistics and Applied Probability*. (Chapman and Hall, London, 1995), Vol. 60.
40. G. G. Walter, *Wavelets and Other Orthogonal Systems with Applications* (CRC Press, Boca Raton, FL, 1994).
41. L. G. Xue, "Approximation Rates of the Error Distribution of Wavelet Estimators of a Density Function under Censorship", J. Statist. Plann. Inference **118** (1–2), 167–183 (2004).